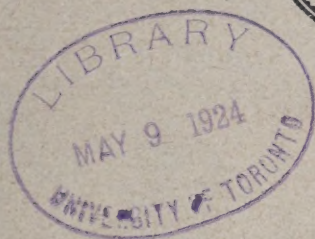


BULLETIN OF THE EXTENSION DIVISION, INDIANA UNIVERSITY

VOL. VIII

BLOOMINGTON, IND.

No. 11



Tenth Conference on Educational Measurements

JULY, 1923

Entered as second-class mail matter, October 15, 1915, at the post-office at Bloomington, Indiana, under the Act of August 24, 1912. Published monthly by Indiana University, from the University Office, Bloomington, Indiana.

**For Sale by the University Bookstore, Bloomington, Ind.
Price 50 cents**

A limited number of copies of this bulletin will be distributed free of charge to citizens of Indiana.

TENTH ANNUAL CONFERENCE
ON
EDUCATIONAL
MEASUREMENTS



Held at Indiana University, Bloomington, Ind., Friday
and Saturday, April 20 and 21, 1923

PUBLISHED BY
THE EXTENSION DIVISION OF INDIANA UNIVERSITY

1923

Contents

	PAGE
THE WORK OF THE BUREAU OF COÖPERATIVE RESEARCH, INDIANA UNIVERSITY. By Henry Lester Smith, Dean of the School of Education, Indiana University.....	3
THE EFFECT OF CERTAIN KINDS OF DRILL EXERCISES ON COMPREHENSION. By Grover H. Alderman, Associate Professor of Elementary Education, Indiana University.....	12
HOW TO INTERPRET AND MAKE USE OF MENTAL TESTS. By Herman H. Young, Associate Professor of Psychology, Indiana University.....	26
IMPROVING INSTRUCTION THRU EDUCATIONAL MEASUREMENT. By Frank W. Ballou, Superintendent of Schools, Washington, D.C.....	44
SOME DANGERS OF SCIENTIFIC MEASUREMENT TO BE AVOIDED (Abstract). By Frank W. Ballou.....	82
EDUCATION AND OUR RESPONSIBILITY FOR ITS IMPROVEMENT. By Frank W. Ballou.....	94
SELLING EDUCATIONAL RESEARCH TO A LARGE CITY SCHOOL SYSTEM. By Robert H. Lane, Assistant Superintendent of Schools, Los Angeles, California.....	102
THE USE OF EDUCATIONAL RESEARCH AND EXPERIMENTATION IN SCHOOL ADMINISTRATION. By Ellis U. Graff, Superintendent of Indianapolis Schools.....	109
THE USE OF EDUCATIONAL RESEARCH AND EXPERIMENTATION IN SCHOOL ADMINISTRATION. By A. C. Senour, Director of Department of Measurements, East Chicago (Ind.) Schools.....	117

The Work of the Bureau of Coöperative Research, Indiana University

HENRY LESTER SMITH, *Dean of the School of Education, Indiana University*

THE work of the Bureau of Coöperative Research, Indiana University, for the current year is organized along nine lines of activity.

In describing these various lines I have felt free in three or four instances to use the wording of the person directly responsible for the study. Particularly is this true in the studies in Latin, methods of teaching reading, tests in home economics, and nursing.

The first of these lines is that of coöperation with educational foundations in carrying on research work. At present two studies are being conducted in an effort to further the program of investigation recently inaugurated by the American Classical League. These two studies have to do with the percentage of the most commonly used words in the English language that are derived from Latin. One person is working on the first 5,000 words of the Thorndike list of the most commonly used words and another is working on the second 5,000 words of the same list.

The only scientific way of determining such a Latin vocabulary is, first, by a word count of the English words used most often in the most widely read reading matter, and, secondly, by an etymological analysis of the English words in the list so obtained.

Dr. Thorndike is now determining the 25,000 English words most frequently used. With his coöperation the committee in charge of the classical investigation will have at their disposal his list for etymological analysis as a basis for determining which Latin words are most important in interpreting the largest number of English words.

Dr. Thorndike's English word count had already progressed far enough a year ago last fall that he was enabled to publish a list of the 10,000 English words most often used in printed and written discourse. The list of 10,000 words already published is to be found in the *Teacher's Word Book* published by Teachers College, Columbia University, 1921.

The scope of the word count used in determining this list of 10,000 words may be most briefly stated by quoting the first paragraph from the instructions for using the *Teacher's Word Book*. I quote as follows: "The *Teacher's Word Book* is an alphabetical list of the 10,000 words which are found to occur most widely in a count of about 625,000 words from literature for children; about 3,000,000 words from the Bible and English classics; about 300,000 words from elementary school textbooks; about 50,000 words from books about farming, cooking, sewing, the trades, and the like; about 90,000 words from the daily newspapers; and about 500,000 words from correspondence. Forty-one different sources were used." This totals about 4,565,000 words in the count upon which the list of 10,000 words is based.

Each of the words in the list of 10,000 has been given an index number weighted to show how frequently it occurs and how widely it is distributed in the forty-one sources upon which the count is based, when compared with the frequency of occurrence and latitude of distribution of the other words in the list.

With this list of 10,000 words as a basis, it is possible to form a list of Latin words that are most important in interpreting the largest number of English words derived from Latin that are in common use. To form this Latin list it is necessary to find the Latin source word for each of the English words derived from the Latin. To determine the place of each of the Latin words in the list in order of relative importance, it is necessary to compare the number of English derivatives of each of the Latin words and also the sums of the index numbers of the English derivatives of each of the Latin words. Thus the Latin word *pars* whose English derivative *part* has an index number of 145 would rank in the Latin list some distance ahead of *persona* whose derivative *person* has an index number of 76.

Latin teachers will naturally be interested to know how many of the Latin words in this Latin list are also to be found in Lodge's list of 2,000 words most important to the traditional course in secondary school Latin. It is proposed to indicate the Latin words that are common to both lists in the bulletin which is to embody the results of this study.

In regard to the 1,000 English words ranking highest by the index numbers in Thorndike's English word list, it is now possible to give some interesting figures. Of the first 500 words ranking highest by index numbers, 82, or 17 per cent, are Latin derivatives. Of the 82 Latin source words, 49, or approximately 60 per cent, are also included in Lodge's list of 2,000 Latin words. Of the second 500 words, 204, or 41 per cent, are derived from Latin. Of the Latin source words of these 204 derivatives, 103, or approximately 51 per cent, are also included in Lodge's list. Summing up for the first thousand, 286 words are derived from the Latin, or approximately 29 per cent. Of these 286 Latin source words, 152, or 53 per cent, are also included in Lodge's list.

As I stated above, two persons are at present engaged in making a study of the Thorndike word list along the lines previously mentioned. The School of Education and the Department of Latin here at Indiana University plan to issue a bulletin in coöperation containing the results of the study. As at present planned, the bulletin is to consist of two parts. Part I will contain three lists. List 1 in Part I will consist of the list of 10,000 English words alphabetically arranged just as they are in Thorndike's *Teacher's Word Book* but, in addition to a second column after the column of words on each page in which index numbers of the words are given, there will be two columns more. In the third column etymological history of the word will be indicated. In the fourth column the Latin or Greek source word will be given when any occurs. List 2 in Part I will contain in the first column the list of Latin source words alphabetically arranged; in Column 2 the sum of the frequency index numbers of the English derivatives of each Latin source word; in Column 3, a number indicating the rank of the Latin source words as compared with

the other Latin source words on the basis of the sums of the frequencies of index numbers of the English derivatives of each; in Column 4, a mark indicating whether, if it happen to be common to Lodge's list of 2,000, it is important in the study of Caesar, or of Cicero or of Vergil; in Column 5, a list of the English derivatives of each of the Latin source words.

List 3 in Part I will contain in Column 1 the list of Greek source words alphabetically arranged; in Column 2, the sum of the frequency index numbers of the English derivatives of each; in Column 3, a number indicating the rank of each of the Greek words in the list; in Column 4, the Latin derivatives of each of the Greek words which are to be found in the Latin list previously given, when there are any such; and in Column 5, a list of the English derivatives of each Greek word.

Part II of the bulletin as at present planned will consist of three subdivisions. Subdivision 1 is to contain a statement of the per cent value of Greek, of Latin, and of each other important source language in giving increased understanding of English as disclosed by comparing (a) the sums of the frequencies or index numbers of the English words derived from each of the languages considered; and (b) the actual numbers of source words of English derivatives in each language.

Subdivision 2 in Part II is to contain a statement of the value of the mastery of Lodge's list of 2,000 words in giving an understanding of the Latin element in the English language. Subdivision 3 in Part II is to contain graphs showing the relative standing of the English, French, Latin, and Greek elements in Thorndike's list, taking the words in order of frequency index numbers by units of 100 in case of the first 500 words and by units of 500 in case of the remaining words of the first 5,000 and perhaps by units of 1,000 words in the case of the second 5,000.

The second line of work that is being undertaken this year by the Bureau aims to determine better procedure in instruction. A somewhat elaborate study along this line in its application to the field of silent reading has been undertaken this year by Grover H. Alderman of the faculty of the School of Education. A detailed report of this study will be made by Mr. Alderman this afternoon. The aim of his study is to measure the relative values of certain types of drill exercises for comprehension in silent reading. Many of you are familiar not only with the aim but with the method of the study, since your schools have been coöperating in the work. For the benefit of those who are not familiar with the plan, the following detailed information is given concerning the method used in the experiment:

"In order to determine what types of exercises to use, a very careful and critical examination of all the experimental studies now published was made in order to note the remedial exercises recommended by the various authorities. Most of the exercises recommended by the various authorities group themselves around a limited number of outstanding types of remedial exercises. For this reason, the three following types of exercises were chosen for experimentation:

1. Improving comprehension ability by increasing the child's reading vocabulary. This was done by giving synonyms and idiomatic expressions plus the proper use of these words in sentences. A notebook of all new words acquired by the child was kept.
2. Improving comprehension ability by means of improving the child's ability not only to organize the important thoughts of the paragraph around a central unit but also to organize the central units of the entire selection in a logical order.
3. Improving comprehension ability by means of improving the child's ability to remember central paragraph thoughts. This was done by means of rigorous testing over materials which had been read not only that day but material which had been read from two to three weeks past.

Practice periods were given 30 minutes each day thruout a period of six weeks. Comprehension was tested at the end of the third and sixth week by means of the Thorndike-McCall Silent Reading Scale.

As a means of determining whether speed might result as a by-product, the Courtis Silent Reading Test was given at the beginning and at the end of the experiment in the schools of Indianapolis.

There were 6 schools, 75 teachers, and 3,201 pupils in grades from 4 to 8 coöperating. Of these, 1,010 were drilled on Type 1; 1,208 on Type 2; 983 on Type 3.

A second study in methods of teaching is this year being undertaken by the Bureau of Coöperative Research in the field of primary reading. In the practice school in the grades maintained jointly by the University and by the public schools, some special methods of presenting the subject of primary reading are being tried out. An effort was made last year and will also probably be continued this year to determine the result of these methods in so far as standardized reading tests will reveal such results.

A third line of work that the Bureau has been engaged in this year is in the field of developing educational tests. One of these, the Van-nest History Test, was finally revised this year and reprinted in a more permanent form than formerly. In the revision, standards of achievement are included.

A second standardized test which the Bureau has been working on this year is the development of a measuring scale in foods. The initiative in this study was taken by Miss Florance Beeson King, of the Department of Home Economics, and was reported on a year ago at the educational measurements conference. A brief summary of this study is as follows:

"Tests were started in the fall of 1922, at which time there were no tests in foods published.

Purpose: 1. To suggest the broader and more desirable subject-matter which should be included in foods courses and to standardize this material to some extent.

2. To set standards of attainments to make possible the comparison of individual pupils, sections, or school systems in their food work.

3. To enable the teacher to measure the efficiency of her teaching.

4. To enlighten the teacher as to the pupil's ability to handle subject-matter and their difficulties.

Method of Procedure: 1. To find subject-matter for tests.

a. Analyzed for content six of the nine textbooks which were found by the University of Chicago survey to be used by 90 per cent of teachers of the United States and three others published since then.

b. Consulted articles on home economics in various educational magazines.

c. Consulted a number of courses of study.

d. Discussed this subject-matter with several teachers of foods in elementary and secondary schools, several university instructors in foods, and one in vocational home-making.

2. Arrangement of subject-matter.

a. Design of test. Simple checking test, with method of right or wrong scoring, covering material involving information and judgment.

3. Preliminary test.

a. Given to group of instructors. All those which were ambiguous or in which there was any difference of opinion thrown out.

b. Given to 100 pupils of a junior and senior high school. All found to be unscorable or otherwise unsatisfactory were eliminated. Arranged in order of increasing difficulty for that group.

c. Given to a group of 50 pupils, 25 answering one set arranged in order of increasing difficulty, 25 answering a set arranged in increasing difficulty within associated groups. First set most satisfactory.

d. Given with above arrangement to about 5,000 pupils from sixth to twelfth grades in Illinois, Nebraska, Ohio, Michigan, Arkansas, and Indiana.

Data from (d) are now being compiled. They will be used to give a score value to each exercise, so that they may be arranged into a tentative scale.

We also hope to form tentative standards of attainment for 7A and 7B and 8A and 8B."

The beginning steps were also taken, looking toward the development of a scale of measurement in the field of mechanical drawing since this subject is offered in the secondary schools in the state of Indiana. A secondary purpose in this study is to investigate the subject-matter in the mechanical drawing curricula in Indiana and to determine the methods that are used in teaching the subject in the high schools of the state. On account of the absence of Mr. Treat, who had this study in charge, little progress has been made on it during the current year.

Miss Florence Blazier, of the Department of Home Economics, of Indiana University, has undertaken a research study for the Bureau of Coöperative Research in a fourth field, namely, that of nursing. This study has been undertaken with the purpose in mind of analyzing this field as an occupation for girls.

The situation has been studied with a view to furnishing information to two groups of people: (1) the vocational counsellor who wishes to advise girls as to their future work; (2) the high school girl who is contemplating entering a school for nurses.

The method of procedure has been as follows:

1. A critical study of all available literature on the subject of nursing education.
2. A study of the extent and nature of the field, by visiting typical hospitals, training schools, public health centers, etc.
3. Study of the results of a questionnaire sent to 1,100 graduate nurses in Indiana, who are engaged in all types of nursing, such as private duty, institutional, and public health.

The questionnaire deals with such points as educational qualifications, including graduate study, age, salaries, with reasons for unemployment and statements of opinions as to the physical and personal qualities necessary for success.

4. Study of the judgments of leaders in the field of nursing who have been consulted in evaluating work of certain items in the questionnaire.

In carrying on this study, it has been very gratifying to find that superintendents of nurses in some of the largest hospitals, heads of visiting nurse associations, and state nursing officials have granted us their fullest coöperation.

This year completes the gathering of data for Professor W. W. Black's study in Handwriting, carried on under the auspices of the Bureau of Coöperative Research.

In May, 1917, writing samples from the 2A grades of 40 Indiana cities were sent to the Bureau. Since that date, samples have been received each year from all or nearly all of the same cities. The Bureau now has samples from grades 2A to 7A. This year samples will be collected from the 8A grades. It is believed that when this material is worked up it should furnish valuable information on a number of points.

1. A study of the handwriting of a large number of individual pupils year by year thru the seven years should give valuable information on pupils' rate of progress in quality and speed and the points of acceleration and slump, if any.
2. A study will be made of the question of when to discontinue formal lessons in writing.
3. It is the intention to make a comparative study of the handwriting of pupils from cities which employ supervisors or special teachers of writing with those from cities in which writing is taught by the regular room teacher.
4. If one or a few of the cities should stand far above the others, a study will be made of the plan and method used in the schools of these cities.
5. The study should give at least a fair estimate of how well writing is taught in Indiana cities in comparison with the standards set for the whole country.

A sixth line of work which the Bureau is undertaking this year is that of assisting individuals and organizations in the state who wish

help along various lines in which the Bureau is especially prepared to render help. Several superintendents, during the year, have written in to the Bureau for bibliographies or for information concerning standardized tests in the fields of measurement of achievement and measurement of intelligence. We are engaged at the present time in working up a bibliography of school surveys. This is being developed in its present form at the request of one of the libraries in the state. We hope later to get it in shape for publication and general distribution.

A third method of coöperation with individuals and institutions out in the state as carried on this year is illustrated by the coöperative study that Indiana University is making for the State Department of Public Instruction. This is a study of the actual combinations of subjects that high school teachers are this year teaching in our own state. The purpose of the study is that it may serve as a basis for some of the constructive recommendations that the State Department of Public Instruction will make to the State Board of Education touching the administration of the new teachers' license law. It was thought that definite concrete information as to the actual present-day demands in our state along these lines would be suggestive in the development of the State Board of Education policy of licensing high school teachers in the future. It is hoped that this brief study may be published soon.

Another field in which the Bureau has been actively engaged this year is that of giving mental tests. Mr. Z. M. Smith, State Director of Vocational Education in Indiana, is conducting this study under the direction of the Bureau of Coöperative Research. The Indiana University mental survey test has been given in connection with this study to 1,500 boys in vocational agricultural schools in Indiana and to an equal number of non-vocational boys in the same schools and of the same school grades as the vocational boys. The same test has been given also in the states of Ohio, New York, Minnesota, South Dakota, California, Iowa, Illinois, and Kansas. The plan was to give the test to a minimum of 100 boys in vocational agricultural schools in each of these states and to an equal number of non-vocational boys in these states. The idea is to use these results in an effort to determine any group differences in mental ability as far as this test is able to reveal differences in native capacity. This study will ultimately serve as a portion of a larger study being undertaken by Mr. Smith entitled "The Work of the Teacher of Vocational Agriculture". This topic he hopes to complete in partial fulfilment of the requirements for the degree of doctor of philosophy.

A further service that the Bureau of Coöperative Research has continued to render this year as it had been doing in previous years is to prepare and keep available an exhibit of educational tests. We now have 282 tests.

The most ambitious study that the Bureau has undertaken this year is that of compiling a bibliography of educational tests. This has been a tedious and a long-drawn-out task. It seemed, however, that there was need for such a study. I have understood only recently that a somewhat similar study is being undertaken by the Bureau of Coöperative

Research at Ohio State University for the National Bureau of Education at Washington, D.C. Our publication, however, is intended to be a little more elaborate than the proposed bibliography to be published by the Bureau of Education. We have the material collected and the data in all cases will cover the following: (1) title of test, (2) author of the test, (3) date of publication of the test, (4) purpose of the test, (5) range of grades covered by the test, (6) time necessary to give the test, (7) time necessary to score the test, (8) a statement as to whether the test has been standardized or not and if so, for what time of the year the standards apply, (9) a brief description of the test. This description is intended to give a fairly adequate idea of the test—for identification purposes and enough to serve as a basis for a general notion of the test.

Following this description is a statement as to the publishers of the test, together with the price at which the test is quoted. Thus far we have listed in this bibliography 301 different educational tests. A chart has been drawn showing the distribution of these tests according to the years in which they were published. The publication of standardized tests and scales of measurement started in this country in 1908 and gradually increased up thru the year 1913, after which there was a very rapid and continuous increase. Our records show that 37 different new standardized achievement tests were published in 1922. Our record also indicates that thus far in the year 1923 eight such tests have come from the press. These numbers relating to the number of tests published in the respective years do not include a list of 65 tests listed as miscellaneous and concerning which the date of publication is not yet available in our records.

Another graph of some interest is made in the form of a tree with its branches [shown with the lecture]. Each branch represents the date of the first test in a particular field; thus, there is one branch to indicate the beginning of the test movement in arithmetic, another branch to reveal similar information for Latin, another one for language, etc. This graph reveals the various subjects of the curriculum that the measurement movement has invaded together with the dates when tests first appeared in these subjects, as follows:

Arithmetic	1908	Grammar	1918
Handwriting	1910	Geometry	1918
Composition	1912	Biology	1918
Drawing	1913	Chemistry	1918
Physical Training	1913	Home Economics	1919
Reading, Spelling	1914	Vocational	1919
Geography	1914	Botany	1919
History	1914	General Science	1919
Algebra	1914	Spanish	1919
English Vocabulary	1915	French	1919
Latin	1915	Journalism	1920
Physics	1916	Music	1920
Language	1917	Poetry Judging	1921
German	1917	Physiology	1921

According to subjects, the number of these 301 tests distributed themselves as follows:

Reading	40	Physical Training	5
Arithmetic	30	Vocabulary	5
Vocational	24	Drawing	4
Handwriting	24	French	4
Composition	20	General Science	4
History	19	Chemistry	3
Spelling	16	Geometry	3
Geography	14	Music	3
Rating	13	Botany	2
Algebra	10	German	2
Latin	10	Spanish	2
Miscellaneous	10	Biology	1
Grammar	9	Journalism	1
Language	8	Modern Language	1
Home Economics	6	Physiology	1
Physics	6	Poetry Judging	1

There is on exhibit on the table at the entrance of the auditorium a typewritten copy of the material for this bibliography in so far as we have gathered the material. Any visitors are free to examine this manuscript for information concerning any line of tests in which they are interested. It is our hope that this material will be available in printed form by the opening of school this coming fall if not at an earlier date.

The Effect of Certain Kinds of Drill Exercises on Comprehension

GROVER H. ALDERMAN, *Associate Professor of Elementary Education,
Indiana University*

I. INTRODUCTION

My purpose for appearing before you this afternoon is to report to this conference the results of a study in silent reading which has been carried on in various schools of the state of Indiana under the direction of our Bureau of Coöperative Research.

For some time we have felt that one of the big problems with which the average school superintendent or supervisor is confronted is the selection of proper drill exercises in silent reading which will prove beneficial when administered by practical teachers under normal classroom conditions. It is true that during the last five years there has been published a large number of scientific investigations in the field of silent reading, but a critical examination of these studies reveals the fact that while they are of great value to the one interested in research, they are, many times, of little value to the busy superintendent in the field.

This is true, not because these investigations are not scientifically compiled, but because the data are obtained many times as the result of intensive training under the direction of an expert, surrounded by ideal conditions. Before the superintendent can recommend them, therefore, he must first make sure that the investigator had the right objectives in mind, and, second, that the data were not obtained at the expense of some other phase of reading which is of equal importance. My contention on this point can well be illustrated in the case of speed in reading. Peters (8) in 1915 drilled pupils in grades three to six for a period of seven months. At the end of this period he reported a gain of 18.7 per cent, but with a loss in comprehension. C. T. Gray (2) two years later reports that in the case of a single fifth-grade pupil it was possible to make a gain as high as 217 per cent. In commenting on the experiment, however, Gray says "Speed was the aim here at the expense of comprehension if necessary."

We believe with Thorndike (10) when he says "No mental function has been deliberately practiced with an eye to improving it and with proper opportunity for the law of effect to operate without some improvement as a result", but, nevertheless, the average school superintendent is not contented with simply telling his teachers to teach silent reading for a certain period each day, without telling them why and how it should be done. The data, therefore, which will be of most value to the superintendent not only tell him *why* and *how* but also *how much*.

One of the most helpful experiments in the entire field of reading was performed by O'Brien (7) in 1920 in the schools of Illinois. In this experiment O'Brien used 40 grades of 20 different schools, in 9 different

school systems. His aim was to determine how much improvement in speed could be made without any loss in comprehension. After two months of drill, O'Brien presented data to prove that the average teacher under careful supervision should be able to obtain a gain in speed of 31 per cent and at the same time obtain, as a by-product, an improvement in comprehension ability of 1.6 per cent. By using three kinds of exercises, O'Brien was able to tell the effectiveness of each. Not only this but he showed that the possibility of improvement in rate did not cease with the sixth grade as had previously been reported by Gray (3) in the *Eighteenth Year-Book*, but that under normal conditions the greatest possibility for improvement occurred in the eighth grade.

With the hope of performing an experiment in comprehension which would prove as helpful to the teacher as did O'Brien's study in speed, the speaker early in the year undertook the task of selecting three kinds of drill exercises to use in the schools of the state for the purpose of improving comprehension ability in silent reading. Two methods of selecting these exercises seemed plausible. These methods for want of a better name we shall call the negative and the positive method. The negative method is that method in which a large number of children are asked to read a difficult selection and then answer a number of questions over the selection which they have read. Their answers when closely analyzed will indicate the causes for their low scores in comprehension. This method was used by Thorndike (10) in his study of paragraphs and sentences. By examining the numerous answers given, he concluded that the reason so many wrong answers were given are: (1) wrong connections with words singly, (2) over potency or "under potency of the elements, (3) failure to treat the ideas produced by reading as provisional until they are accepted or rejected.

While this method has its merits, the speaker realized that should he attempt to follow it, his investigation would require that the papers be scored by a number of judges in order to eliminate the personal judgment of the speaker in scoring. Also if this method were followed, he would not be able to utilize the large number of scientific investigations which have already been made in the field of silent reading, especially those which have to do with comprehension. For this reason it was deemed wise to make a very thoro examination of all the scientific studies in the field of silent reading which gave emphasis to the subject of comprehension, in order to learn whether there had been any investigations reported which might lend themselves to further experimentation.

1. Need for Vocabulary Training. One does not have to delve very deeply into the literature of silent reading until he is impressed with the fact that, altho approaching it from different angles, all investigators agree that one of the chief causes for low scores in comprehension is the lack of an adequate vocabulary. Thorndike was the first to point this out in his article "Reading as Reasoning" (11). Gray also emphasizes it in his article in the *Twentieth Year-Book* (4) in his description of a seventh-grade boy who ranked low in comprehension because of a limited

vocabulary. As a result of six weeks of training on vocabulary, in which words were studied in context for meaning, with special emphasis placed on prefix, suffix, and stem in the case of the most difficult words, Gray says, "Suffice it to say that the special training given in this connection resulted in a 50 per cent deduction in error and a gain in rate." In comparing the progress of this boy with that of a boy who did not receive the special training, Gray says, "In comprehension, the subject of the experiment made far greater progress than the check pupil, altho at the close of the experiment he was still behind the standard for the class."

Pressey and Pressey (9) in making a rather critical study of the concept of silent reading selected a large reading vocabulary as one of the four factors needed in the reformation of silent reading problems. They say: "Evidently a reading vocabulary is a factor most readily acquired by reading. Measurement of vocabulary the writers feel should be a regular practice in appraising the work of a teacher or a school." References like the one just cited, with many others of equal importance, lead one to believe that one of the most important exercises for increasing comprehension ability is one which has for its aim the increasing of the child's reading and understanding vocabulary.

2. **Need for Training in Organization.** A second type of difficulty which is frequently mentioned by writers dealing with the subject of low scores in comprehension is the lack of power to select central paragraph thoughts and organize them in a logical manner. Most writers fail to recognize the fact that in order to organize properly one must have the proper elements to organize. That children do not possess the ability to select central paragraph thoughts is pointed out by both Germane (1) and Woody (12). Woody gave his silent reading test to 1,245 children for the purpose of seeing how well children were able to pick out central paragraph thoughts. Woody says, "Out of the nine exercises in the test the medium number correctly marked varied from 2.6 in the fourth grade to 5.5 in the eighth grade. This study substantiated Germane's study that 'most pupils in grades five to nine do not know how to summarize the material read but devote most of their time to indiscriminate note taking.'"

Regarding the ability of pupils properly to organize central thoughts in a logical manner Thorndike (11) says "Understanding a spoken or written paragraph is then a matter of habit connecting mental bonds, but these have to be selected from so many others, and given relative weights delicately, and used together in so elaborate an organization, that 'to read' means 'to think' as truly as it does 'to evaluate' or 'to verify'!" Such evidences as those cited by Germane, Woody, and Thorndike led the speaker to select for his second type of exercise some drill work which tended to train the child, first, to select the central paragraph thought, and second, to organize these thoughts in a logical and an orderly manner.

3. **Need for Training in Retention and Reproduction.** Third and last, the speaker found that many writers considered the inability to

retain and reproduce what had been read as a vital factor in successful silent reading. Horn (6) considers that rate, comprehension, organization, and retention are practically on a par so far as successful silent reading is concerned. He points out the fact that special drill work should be devised in order to make the child more proficient in each ability.

One of the most elaborate experiments in retention which has been performed in recent years is reported by Yoakam (13) in the *Twentieth Year-Book*. In this article Yoakam gives data to prove that the ability of the average child to reproduce that which he has read varies from 7.5 per cent in grade four to 20.9 per cent in the eighth grade. In other words, the average child in the elementary school, in the language of central tendencies, is only able to reproduce approximately 10 per cent of the factual material as a result of a single reading. An exercise which aimed to increase this ability was selected for the third type of drill work.

A report of the principal causes underlying low scores in comprehension would not be complete without a brief report of an unpublished study just completed by Dr. George Hilliard (5) of the Western State Normal School at Kalamazoo, Mich. After making a thoro survey of all the educational literature written in the field of silent reading, Hilliard lists six causes for low scores in comprehension tests. These six are: (1) low general intelligence, (2) insufficient vocabulary, (3) lack of organization ability, (4) faulty rate of reading, (5) inability to reproduce, (6) excessive lip movement and articulation. In order to determine the correlation between these six factors and comprehension ability, Hilliard measured the comprehension ability of 161 fifth-grade pupils by the Monroe, the Burgess, the Courtis, and the Thorndike-McCall Scale and correlated these results with (1) intelligence as measured by the Terman Group Test, (2) vocabulary as measured by the Thorndike Visual, (3) organization as measured by Greene's Organization Test, (4) rate as measured by the Courtis tests, (5) and reproduction as measured by the Dutch Homestead Test. In commenting on his results Hilliard says, "Comparisons between the scores on comprehension tests and measures of the six factors given above show a positive relationship between them all." The zero order of correlations gives the order of importance of these six factors as follows: (1) intelligence, (2) vocabulary, (3) organization, (4) rate, (5) reproduction, (6) lip movement.

TABLE I.
(After Hilliard)

TYPE OF TESTS	Monroe		T. McC.		Burgess		Courtis	
	r	P.E.	r	P.E.	r	P.E.	r	P.E.
I. Intelligence Terman Group.	609	085	827	041	906	024	709	69
II. Vocabulary Thorndike, Visual Vo- cabulary	425	.097	.689	071	.599	086	.536	.096
III. Organization Greene.	250	126	.613	084	.549	094	.566	092
IV. Rate Courtis.	711	067	485	103	726	067	.685	076
V. Reproduction Dutch Homestead.	343	119	277	114	615	083	643	079

Note: The number of cases (25) is so small that unless the correlation is five or six times the probable error it has little significance.

Since all will agree that native intelligence does not improve with training, the three kinds of drill work selected by the speaker coincide with the ones selected by Hilliard with the exception of rate. While rate was recognized by us as an important factor in improving comprehension ability, this element was purposely omitted since we felt that the relation of rate to comprehension had been so well solved by O'Brien.

The three types of exercises finally selected for experimentation are:

Type I. Those which aim to enlarge the child's reading and understanding vocabulary.

Type II. Those which aim to improve the child's ability in selecting central paragraph thoughts and organizing them in a logical manner.

Type III. Those which aim to improve the child's ability to retain and reproduce the important points which he has read.

II. THE PROBLEM

The problem, therefore, which was finally selected for experimentation is briefly stated as follows:

1. Under normal classroom conditions is it possible to improve comprehension ability by drilling intensively over a period of six weeks?

2. If comprehension ability can be improved, which kind of drill exercise is the most effective?

III. METHOD

In order that the experiment might be carried on under normal classroom conditions, several schools of the state were invited to co-operate. The response to the invitation sent out by Dean Smith was so cordial that more schools responded than we were able to use, since the expense incurred, in conducting such an experiment, made it prohibitive to use more cases than was necessary. On the other hand, there are always so many unforeseen conditions which arise that make it necessary to include a large number of cases to offset any such variable. In all, there were 3,201 pupils and 75 teachers in grades four to eight in the cities of East Chicago, Frankfort, Lebanon, Mishawaka, Huntington, and Indianapolis who coöperated in this experiment. Of this number, 1,010 children under the direction of 25 teachers were used in the vocabulary experiment, 1,208 pupils under the direction of 25 teachers were used in the organization experiment, and 983 children with 25 teachers were used in the retention experiment. The cause for the unequal number of children in the various types is due to the fact that each school system worked on all three types of exercises. This was done to eliminate any variable which might arise as a result of better teaching or better supervision which might occur in certain cities.

The instructions sent to the teachers were carefully explained in every detail, covering thirteen pages of closely typed material. The aim of the experiment, the method of dividing the class, the method of teaching, and the methods of administering the tests were carefully explained. While time does not permit me to explain in detail the various instructions sent to the teachers, it does seem necessary to explain somewhat briefly the chief points of interest.

The Thorndike-McCall Reading Scale, Form I, was used to divide each class into a drill and a non-drill section. The teacher was advised to divide the class into two divisions of equal reading ability as indicated by the T. Scores on the Thorndike-McCall Scale.

After the division had been made, the teacher was directed to conduct the non-drill group by the same method she had always used with the exception that the recitation in all cases was to be but thirty minutes in length. It was also suggested that she keep the drill and the non-drill group separated during the recitation period with the exception of those days when the Thorndike-McCall tests were to be given. The teachers, as a rule, found it impossible to comply with this request since neither the extra room nor the extra teacher was available.

While no conferences were held nor demonstration lessons taught, as was done by O'Brien in Illinois, each teacher was furnished a five-page mimeographed pamphlet which explained in detail the method to be followed in conducting the drill class. While the instruction varied with each type of exercise, there were a few points in common which applied to all three types. First it was explained to the teacher that the reading material to be used by the drill group should always be of a concrete nature, well organized and easily tested. It was suggested that the teachers use selections on citizenship, health, commerce, geog-

raphy, history, etc. In order that the material would present a real problem in comprehension to the child, it was suggested that the selections be taken from the grade above. It was emphasized that neither fairy tales nor poetry was to be used by the drill group.

The second instruction which was common to all grades had to do with the method of conducting the first day's recitation. During the first half of the first day's lesson each teacher was asked to explain to her class the aim of the experiment. She was also asked to prepare the minds of the children by explaining to them the importance of comprehension, the relation of comprehension to study habits, and the possibility of improving it. It was further explained that there were 75 different classes in the experiment, and that this particular class must do its best since the results were to be published.

The second half of the first lesson was used to introduce the child to the type of drill work he was to use. In the case of Type I, the vocabulary study, the child was asked to read silently a rather difficult selection which was taken from the grade above. After the class had finished reading the selection once, some child was asked to tell in his own words the important thoughts of the selection. As soon as the teacher discovered the child's inability to do this, she pointed out to him that his inability to select central paragraph thoughts was largely due to the fact that he had failed to understand the meaning of certain fundamental words. The teacher then proceeded to explain that during the next six weeks intensive drill work would be given for the purpose of building up each child's reading vocabulary.

The method used for the vocabulary study was similar to that reported by Gray (4) in the *Twentieth Year-Book* of the National Society for the Study of Education. The method in brief is as follows: As soon as the child encountered a word that he did not understand, he looked it up in the dictionary, not for the purpose of learning its definition but rather for the purpose of learning several synonyms and idiomatic expressions which explain its meaning. The word with several synonyms and equivalent idiomatic expressions and an illustrative sentence were then recorded in the child's notebook which he was advised to keep from day to day. For example, instead of writing the definition of the word "indomitable", the child would simply record "indomitable—meaning brave, fearless—idiomatic expression, to look danger in the face—Sentence, He was an indomitable hero."

The method used in Type II was similar to that used in Type I in that the first part of the first lesson was used in explaining the aim of the experiment, the method of approach, etc. In Type II, however, the pupil was led to see that one of the chief causes for low scores in comprehension is due to the fact that children are not trained to pick out the important thoughts of a selection and organize them in a logical way. It was further explained that during the next six weeks the class would be drilled for thirty minutes each day, to see if it would not be possible to improve each child's ability in selecting central paragraph thoughts and organizing them in an orderly and a logical manner.

The method used in teaching organization was briefly as follows:

The children were asked to read the selection rapidly. After the reading was completed, the important thoughts of each paragraph were then selected and written on the board. The children were then permitted to read the selection the second time to correct any wrong impressions which they had gained as a result of the first reading. After this had been done, the children in coöperation with the teacher organized these thoughts in their proper relationship, in the form of a brief, concise outline. These outlines when completed were placed in the children's notebooks in order that they might be used for future review work.

The third type of exercise differed from Type II in that the aim here was to improve comprehension ability by means of constant drill work in the reproduction of important facts taken from the reading lesson. The instructions in Type III provided for the following: The teachers were requested to ask each pupil to read his lesson rapidly, keeping in mind that he was soon to be submitted to a short, rigorous fact examination. As soon as the class had read the selection once, ten short fact questions were given which could be answered in one word. After the test was completed the papers were exchanged and corrections made. After the papers had been corrected and passed back to the original owners, the class was directed to read the lesson the second time, with the hope of finding the correct answer to each question. In order to facilitate matters, the teacher was requested to have the ten questions written on the blackboard under a map so that the questions might be plainly seen by the class when the map was removed.

After the children had been drilled for three weeks by one of the three methods described above, they were given the Thorndike-McCall Reading Scale, Form II, to measure any improvement which had been made in comprehension. After this the children were again drilled for three weeks, and on Friday of the sixth week Form III of the Thorndike-McCall Scale was given to all classes.

The testing program outlined above was followed by all schools with the exception of Indianapolis. Here, on account of a shortage of Form I tests, Form III was given first, followed by Forms II and I. Theoretically this should make no difference in the results as all of these scales are of equal difficulty from a statistical point of view.

In order that a gain in comprehension might not take place at the expense of some other element of equal importance, the teachers were urged to have all reading done as rapidly as possible, provided that it was not encouraged beyond the point of efficiency. In order to make sure that the children did not lose in speed, 500 children from the schools of Indianapolis were given the Curtis Silent Reading Tests, Forms I and II. It was planned that Form I be given on the first day and that Form II be given on the last day of the experiment, but owing to circumstances over which we had no control but three weeks intervened between the administering of the first and second tests.

IV. DATA

The data which I shall present today is a compilation of the returns from only 35 teachers. As has already been explained, there were in all

75 teachers coöperating in this experiment, but on account of spring vacations one school system started with the experiment one week late, and, as a result, the returns from this city were not sent in to the office soon enough to include the data for presentation today. The report today, therefore, will give data on only 1,129 pupils, distributed

TABLE II.—GENERAL INFORMATION.

	Type I.	II.	III.	Total
Teachers	13	13	9	35
Drill	203	210	163	576
Non-Drill.	195	202	156	553
Total	398	412	319	1,129

among the various types as follows: 398 in Type I, 412 in Type II, and 319 in Type III. The classes originally were divided in such a manner that there were an equal number in the drill and the non-drill sections. The inequality shown in Table I is due to the fact that a child's score was not considered unless he took both tests. As a result, the number of pupils in the drill and the non-drill sections vary as is shown in Table II.

As has been previously stated, it was the original plan to report the amount of improvement made after three weeks of drill and again after six weeks of drill. These data, with the data compiled from the Courtis tests, will be tabulated within the next few weeks, but time did not permit the work to be completed for presentation at this conference. The report today gives data on only two tests: Form I, which was given before the drill work began, and Form III, which was given on the last day of the drill period.

TABLE III. —AVERAGE GAIN BY SECTIONS

Form	3	1	Average Gain
Drill	50.21	46.40	3.81
Non-Drill	47.75	46.29	1.46
Difference			2.36

Table III shows the average gain in T. scores for the drill and non-drill groups. The average score for the non-drill group before training was 46.29 while the average score after training was 47.75 or a gain of 1.46. The average score for the non-drill group before training was 46.40 while the average score after training is 50.21, or a gain of 3.81

for the drill group. This table shows that the average grade child, under the teaching he is receiving in these particular schools, gains normally during a period of six weeks 1.46 points on the Thorndike-McCall Scale, while children of approximately equal reading ability gain 3.81 points on the same scale, after having been subjected to intensive drill exercises for thirty minutes each day for a period of six weeks.

While data of the type presented in Table III are gratifying to the one directing the experiment, they are of little value to one interested in supervision. If gains are made the supervisor wants to know, first, which particular teachers are responsible for the gains, and, second, which type of drill work proved most effective. Table IV gives this information.

TABLE IV.—SHOWING SECTIONS WHICH RECEIVED HIGHER SCORE

	I.	II.	III.	Total
Drill.....	7	7	7	21
Non-Drill..	6	6	2	14
Total.....	13	13	9	35

In Type I, 7 of the 13 classes profited by the drill work, while 6 failed to profit. In Type II, the results were similar to those in Type I, but in Type III, while there are but 9 classes reported, it is important to notice that 7 of the 9 sections profit by the drill work. Comparing the drill sections with the non-drill sections, 21 of the drill sections profited by the drill work, while 14 failed to profit.

Since the important element in any teaching is the individual child, the supervisor is always interested in knowing the value of an exercise not only in the language of central tendencies, but also in terms of the individual child.

TABLE V.—NUMBER OF PUPILS GAINING

TYPE	Gain		Loss		Same		Total	
	A	B	A	B	A	B	A	B
I.....	151	133	49	45	12	16	203	195
II..	170	143	32	52	8	7	210	202
III. --	111	105	47	43	5	8	163	156
Total.....	432	381	119	140	25	31	576	553

Table V is read as follows: Of the 203 children who were drilled in Type I exercise, 151 gained, 40 lost, and 12 received the same score in both tests.

TABLE VI.—PERCENTAGE OF PUPILS GAINING

TYPES	I.	II.	III.	Total
Drill	74	80	68	75
Non-Drill	68	71	67	69
Drill over Non-Drill	6	9	1	6

Table V reduced to percentages shows that of the 398 children using drill exercise Type I, 74 per cent of the drill group profited while but 68 per cent of the children, in the non-drill group, increased their scores during the six weeks' period. Using the percentage of pupils as a method of determining the relative value of these three types of exercises, Table V shows Type II to be first with 80 per cent, Type I second with 74 per cent, and Type III third with a score of 68 per cent.

Data showing the number of pupils that gain as a result of any exercise are a more accurate measure of a teaching method than data which show the average gain for each pupil.

TABLE VII.—AVERAGE T. SCORE GAIN PER PUPIL

	I.	II.	III.
Drill	3.13	5.10	2.99
Non-Drill	3.11	3.80	2.78
	.02	1.30	.21

Even by the method of average T. Score gains, Drill Exercise Type II would again receive first rank.

In any kind of drill work, the supervisor is always interested in knowing the type of pupils who are benefited by the exercise. A drill exercise which is of value to the class as a whole may prove a detriment to the pupils in the upper quartile.

TABLE VIII. PERCENTAGE OF GAIN BY QUARTILES

	Type I.	
	Q1	Q1
Drill.....	64	80
Non-Drill..	46	84
	Type II.	
	Q1	Q1
Drill...	61	92
Non-Drill..	51	83
	Type III.	
	Q1	Q1
Drill.....	50	82
Non-Drill.....	50	75

Table VIII shows that while the exercises used in this experiment were of some value to the pupils of the upper quartile, they proved to be of most value to the pupils of the lower quartile. It is interesting to note that from the standpoint of the better readers, as measured by the Thorndike-McCall Scale, the exercises are of value in the order named, while from the standpoint of the slower pupil Type II ranks first, Type III ranks second, with Type I ranking third.

V. CONCLUSIONS

While one hesitates to draw any conclusions before all the data are tabulated, the facts presented this afternoon seem to warrant certain tentative conclusions.

1. Comprehension ability in silent reading as measured by the Thorndike-McCall scale may be improved to a degree equivalent to nearly one semester of work in grades four to eight by careful, systematic drill work covering a period of six consecutive weeks, provided that thirty minutes each day is devoted to this type of work.

2. A teacher who is interested in improving comprehension ability in silent reading as measured by the Thorndike-McCall Scale would be justified in using any one of the exercises used in this experiment.

3. The relative value of the exercises used in this experiment vary with the reading ability of the pupil. Type II, or the organization exercise, is of most value to the average pupil, and also to the pupils in the lower quartile, while Type I, or the vocabulary study, is of most value to the pupil in the upper quartile.

VI. AN EVALUATION

Anyone who has had any experience in the field of experimental education is aware of the fact that the results of this experiment are

by no means startling. There are several facts, however, which one should keep in mind while attempting to estimate its worth. This experiment was performed in six of the best school systems of the state. In these schools the subject of silent reading had been emphasized throughout the entire school year. In order to make sure that the experiment would not fall into the hands of inexperienced teachers, each superintendent was requested to permit only those teachers to participate who were especially interested in the teaching of silent reading. Had this experiment been carefully supervised in schools in which the subject of silent reading had not been stressed, more gains might have resulted.

Altho the data showed that the drill work in organization proved the most effective means of improving comprehension, it must be remembered that these conclusions were made after comparing the scores made by the drill group with those of the non-drill group. In our instructions to teachers it was suggested that the drill group recite in a separate room. The teacher as a rule found it impossible to comply with the request. With this thought in mind it is interesting to note that the work in organization, an exercise which was performed largely as an individual problem, proved to be of most value to the average pupil.

Again while exercises one and three proved to be of less value than did exercise two, the results might have been different had a finer measuring scale been used. A close examination of the steps on the Thorndike-McCall Scale reveals the fact that each unit on the scale is equal to an age difference of from three to six months. While the results of the six weeks of practice may have resulted in considerable gain, this gain may not have been equal to six months of regular school work. This study indicates that we need further experimentation using finer measuring scales.

BIBLIOGRAPHY

1. Germane, C. E. Outlining and Summarizing Compared with Re-reading the Same Article. *Twentieth Year-Book*. Part II, p. 109.
2. Gray, C. T. Type of Reading Ability as Exhibited through Tests and Late Experiments. *Sup. Ed. Mon.*, Vol. I, No. 5, Aug., 1915, University of Chicago.
3. Gray, W. S. *Eighteenth Year-Book of the N.S.S.E.* Pt. II, p. 28.
4. Gray, W. S. *Twentieth Year-Book of the N.S.S.E.* Pt. II, p. 49, Public School Publishing Co., Bloomington, Ill.
5. Hilliard, George H. Probable Types of Difficulties Underlying Low Scores in Comprehension Tests. (An unpublished doctor's dissertation, University of Iowa), 1922.
6. Horn, E. A Constructive Program in Silent Reading. *Jr. of Ed. Research*, May, 1921.
7. O'Brien, John A. *Silent Reading*. The Macmillan Co., 1921.
8. Peters, C. C. The Influence of Speed Drills upon the Rate and Effectiveness of Silent Reading. *Jr. of Ed. Psy.*, Vol. VIII, 350-66.

9. Pressey, L. W., and Pressey, S. L. A Critical Study of the Concept of Silent Reading. *Jr. of Ed. Psy.*, Vol. XII, pp. 25-31.
10. Thorndike, E. L. *Educational Psychology*. Vol. II, p. 151, Teachers College, Columbia University, New York City.
11. Thorndike, E. L. Reading as Reasoning. *Jr. of Ed. Psy.*, Vol. VIII, p. 323.
12. Woody, C. W. Measuring of a New Phase of Reading. Unpublished manuscript.
13. Yoakam, G. A. The Effect of a Single Reading. *Twentieth Year-Book of the N.S.S.E.*, Pt. II, p. 90.

How to Interpret and Make Use of Mental Tests

HERMAN H. YOUNG, *Associate Professor of Clinical Psychology, Indiana University*

A MENTAL test is merely a specially constructed arrangement of mental stunts to be uniformly presented to all persons selected for testing. In psychological phrasology a mental test is a stimulus or series of stimuli presented to an individual or group of individuals to elicit a verbal response or motor reaction under such conditions that a permanent record thereof is made, either by the individual who is being tested or by the one doing the testing. Since we are now interested primarily in group mental tests it should be recognized that a group mental test is a series of stimuli presented by a predetermined and uniform method to groups of individuals to elicit a written response from each member of the group. Most group mental tests are constructed so that the time and energy required in writing the response is very small. An individual's score is largely the result of his mental reaction to the test and not the result of time and energy consumed in the physical manipulation of the pencil in writing the response.

Vitiating, unstandardized conditions which vary from individual to individual and from school to school and which, when present, seriously modify the scores of the individuals or group affected, may be classified under three headings: (1) physical conditions of the person being tested, (2) failure on the part of the tester to present the tests in accordance with the standardized instructions, and (3) particular nature of the physical surroundings under which the test is presented.

Physical conditions which may affect the score of a person are of two types, temporary or permanent, and include such conditions as defective vision, defective hearing, physical deformities, fatigue, malnutrition, lack of sleep, headache, and illness during the test.

Faulty presentation of the tests includes such things as the following: The tester does not talk loud enough, he talks too rapidly, he alters instructions in accordance with his own judgment, he gives extra suggestions, his manner of conduct is such that the children look upon him as a curiosity, he rushes nervously about and gets everybody's nerves on edge, he forgets part of the instructions, he disregards time limits, he unnecessarily scolds a child and is bossy.

Physical surroundings during the test include such things as breaking pencils, radiators discharging steam, honking of automobile horns, the ringing of bells and whistling of passing trains while verbal instructions are being given, poor light, testing thru recess period or after closing time, or while other children are at recess, principal or teacher

boisterously and possibly unnecessarily scolding a child, some curious individual making a nuisance of himself by passing up and down the aisle while the children are at work, and lack of sufficient desk space to write comfortably.

Besides these special conditions which affect the scores of only that particular individual or that particular group which happens to be the victim of such unfortunate circumstances, there are considerations and features of test construction and interpretation which affect all persons tested. Anything at all can be presented as a mental test, but not all things that might be whipped into test form belong in so-called mental tests. The fundamental criteria can be stated in some such a manner as this: A mental test should present an extensive variety of stimuli of such nature that the examinee has never before encountered them in that form. Ideally, the variety of stimuli should be extensive enough to make demands upon every type and phase of mental activity, in order, on the one hand, to allow the examinee to display his mentally strong points, and, on the other hand, to cause him to reveal his weak points. As the presenting of stimuli in the same form in which they have been experienced previously emphasizes memory, it is necessary to present them in a manner new to the individual, else only memory, that one phase of mental activity, will be tested. As the presenting of entirely new stimuli, no part of which has ever been experienced previously, calls forth nothing but a guess or chance reaction, if any at all, it is necessary to present stimuli that are new in no respect except the particular setting in which they are presented in the mental test, i.e. they must tap old experiences from a new angle.

It is impossible to devise any group test which will meet these ideal demands, because no two people ever have identically the same environment and experiences. This does not, however, preclude the possibility of constructing a series of tests which for all practical purposes are equally fair to the different members of selected groups. On account of special environmental settings one person will be enabled to answer certain questions from memory, and will have little or no experience to assist him in answering certain other questions. Varying from individual to individual, never two alike, different memory successes will be made possible and different failures inevitable according to the peculiar environmental and educational background which each happens to have had. Thus each individual will increase his score by passing certain tests largely from memory while others will fail on these same tests just because their environment did not give them the necessary experiences. Stated inversely, each will fail to score on certain tests because his environment has not given him the necessary experiences, while others will pass these same tests from memory.

With so many variable factors actively involved in producing mental test scores our interpretations and use of them must be made cautiously and only in full recognition of our limitation in so doing.

Most of the conditions already enumerated as capable of seriously affecting an individual's score can do so only by preventing it from being as large as it would be under normal conditions. It is obvious

that an individual may fall very low on a test because of some temporary physical condition, because of some fault of the examiner, or because of the surroundings under which he is required to take the test, or because of all of these. Even tho these are only occasional causes of failure, they may have serious bearings in those particular cases who unfortunately are the victims of these special conditions, and they must ever be kept in mind when passing judgment upon those making low scores. Because these vitiating influences may strike we know not where, and obviously not according to any rule, it is impossible to formulate any general method of handling them except to be on the alert for their presence and to make due allowances for them when found.

Suppose there were no such disturbing influences as those mentioned above, that every person were physically perfect, that the tests were presented exactly as they should be, and that the testing conditions were ideal, how should we interpret and use such test results? Before answering this question it is necessary to consider the various factors at work in producing test scores. Since it is impossible to devise a mental test specifically applicable to more than one person, the fundamental criteria underlying the construction of group mental tests must be reckoned with. The most important criterion is: that the test should present, in a manner new to every member of the group, a wide range of different types of stimuli which have come within the realm of their previous experience.

We have already noted why a wide range of stimuli is necessary and why they must be based upon past experiences. Our next task is to attempt to determine the significance of scores resulting from such test construction. This can be done only thru use of carefully defined terms.

The term "score" has a definite meaning and is universally understood as the number of credit points which an individual receives for his answers. The measures of central tendency, average, median and mode, and those of variability, minimum and maximum, mean deviation, quartile deviation, and standard deviation, and the measures of relation or correlation are universally standardized in the statistical treatment and interpretation of scores.

The terms "mental ability" and "intelligence" are used in so many different ways that they have confused rather than clarified, have hindered rather than facilitated the proper interpretation and use of mental tests. This has occurred because, purporting to be basic and elemental concepts, their use has discouraged and actually delayed further analysis. This unanalyzed usage has permitted and encouraged the belief that test scores are as direct and reliable a measure of one's mental ability and intelligence as pounds are of one's weight. This might be true if by mental ability and intelligence is meant the individual's ability to do things at the very time he is taking the test, just as by weight we mean how much he weighs at the time he is on the scales. But it can hardly be true if by mental ability and intelligence we mean the

Methods of determining and interpreting these are given in detail in Harold O. Rensis' *Statistical Methods Applied to Education*.

individual's innate capacity for mental development, unless by weight is meant an individual's innate capacity for physical development.

Only thru the use of unambiguous terms will clarity be assured and progress be made in analyzing test results. For this reason the two terms, "mental ability" and "intelligence", have no constructive value. Other terms with definite and entirely distinct meanings are necessary to permit and force a clear distinction between the things an individual does and the significance of those things. Whenever an individual does anything at all, be it the winking of an eye, the signing of an international treaty, or the taking of a mental test, he is performing an act or series of acts. The score of an individual on a group test is only a numerical statement of the credit value of his performance, i.e. of what he actually did on that particular test. As an end in themselves this performance and its score have no value. Properly handled, they may serve as a basis for predicting a child's probable development in the future. The only way in which future developments may be predicted is in terms of past accomplishments and present performances.

It would therefore seem advisable to employ the term "performance" as Witmer defines it, as the doing of anything whatever, be it very simple or very complex. The taking of a mental test is a performance which is evaluated in terms of credit points, called scores. It is not possible in terms of score alone to predict the probable future standing of an individual. This situation demands the introduction of a term which clearly distinguishes a performance from the significance of that performance, which distinguishes an observed or determined fact from the interpretation thereof, and which definitely emphasizes the difference between the thing accomplished and the predictive value of that accomplishment. It is undoubtedly the failure to make this distinction that has caused more mistakes and more criticisms in mental testing than any other one thing.

Tests should be employed to serve two purposes; first to determine what an individual actually does at the time he is tested, just as scales are used to determine what a person actually weighs at the time he is on the scales; and, second, as a point of departure in predicting what may be expected of him in the future. Under favorable conditions an individual is likely to continue in his mental and physical development according to certain fairly well-established principles. It should therefore be possible to so interpret an individual's present performance that prediction could be made with a fair degree of reliability as to the probable nature and extent of his future performances. The term "developability" is employed here to designate the probable extensiveness and complexity of the highest performance level or upper limits of development for which an individual furnishes prospects.

The term "performance" applies only to what the individual actually does. The term "developability" is used to designate the probable highest level of performances the individual will ever execute under the most favorable circumstances for development.

These definitely defined terms enable a direct attack to be made upon the problem of interpreting mental test results. There are three pos-

sibilities: first, that the score of an individual indicates the extensiveness of his environmental opportunities; second, that it indicates his developability; and third, that it only indicates his level of performance at the time he takes the test. If the first assumption were correct, the individual with the best environmental opportunities would have the widest range of helpful experiences and would make the highest score, the one with the next best environmental opportunities would have the next widest range of helpful experiences and would make the next highest score, and so on for all the individuals down to the one having the poorest environmental opportunities. He would have the narrowest range of helpful experiences and would make the poorest score. If this assumption were correct, scores would be a measure of an individual's past environmental opportunities.

If the second assumption were correct, the individual with the highest developability would make the highest score, the one nearest him in this respect would make the next highest score, and so on down to the one with the lowest developability. He would make the lowest score. If this assumption were correct, scores would be a measure of the complexity and extensiveness of an individual's developability.

If the third assumption were correct, the individual with the highest developability would probably make the highest score only in case he had had the advantages of more helpful environmental influences than any other individual. It is also probable that the individual with the largest number of helpful environmental influences might make a better score than others of superior developability but not favored with such a large number of helpful environmental influences. If this assumption were correct, scores could only be a measure of what an individual actually does on the test.

When viewed in this perspective there is no doubt as to which of these three possibilities corresponds closest to facts. It is obviously preposterous to assume that mental tests measure the extensiveness of an individual's past environmental influences, and it is just as preposterous to assume that they measure his developability. The indispensable value of experience has already been indicated. Without it human beings would be as incapable of responding to mental tests as inanimate objects are. In fact that is the case when they are given tests no part of which has ever come within the realm of their experience.

Two, and only two, measurable things determine what an individual will do on a mental test. The first is the nature and extent of his past environmental influences; the second is the effect produced upon or within him by these environmental influences. It is reasonable to expect that the effect which an environment can have on a person will vary with the length of time it has had opportunity to act upon him.

The most probable effect which environment will have on him in the future can be reliably predicted only in terms of the effect which his past environment has had upon him. Prediction can therefore be made only in terms of the relative effectiveness of past environmental influences upon the individual. The general nature and extent of the environmental influences to which an individual has been subjected can be

satisfactorily determined without much difficulty. There are two entirely separate and distinct phases of environmental influences which must always be considered. They are the particular nature or type of environment to which the individual has been exposed and the extent or length of time of the exposure. Therefore the particular effect which an environment can be expected to have upon an individual will depend upon its specific nature and the length of time he has been subjected to it.

The important facts of an individual's past life are the nature of the environment or environments to which he has been exposed and the length of time he has been exposed to each. It would seem reasonable to expect that the effectiveness of a given environment upon a given individual would vary directly with the length of time he has been exposed to it. This would seem to be especially applicable with children reared in the same community, because so far as environmental influences and experiences are concerned all children start from the same point, or zero. Each year as the length of environmental exposure increases they do more things and things more complex. Thru use of scientific methods of investigation it is possible to determine about how many and what kinds of things children of each age do. In this way, and in this way only, can the standard of accomplishment for each age be determined. Mental tests are a modern device for determining what children do under specified conditions. It is extremely important to note that tests do only this and nothing more. After a large number of children have been tested on a particular test it is possible to treat the scores statistically and find the average number of credit points which children of each age earn. This average is then an established standard of what children of each of the various ages actually do under these standardized conditions, and may be taken as the average performance level for the respective ages. Thus the average score for six-year-old children establishes a standard of performance. Because it is the average of the performance of six-year-old children it constitutes the six-year-old performance level. In the same way a performance level may be established for each age. Whenever a child earns the appropriate number of credit points on a test to give him the score that is the average for any given age, he may be referred to as having the performance level of that particular age. Thus to say that a child has a six-year-performance level means one thing and only one thing, i.e. that he performed the test in such a manner that he received the same number of credit points as the average for six-year-old children. The difference between saying this and saying he has a mental age of six, implying thereby that he is like a six-year-old child, is greater than the difference between day and night. Only six-year-old children can be like six-year-old children. This does not mean any more about the resemblance of the children in other respects than to say two men are alike because both drive Fords. If nothing else besides a child's score is known, prediction as to his future is more impossible and ridiculous than prediction as to the future of the Ford drivers.

An uninterpreted mental test score is useless. A misinterpreted

mental test score is vicious. Reliable interpretation and use of test scores can be made only by giving due weight to the known and measurable factors involved in their production. The suggested interpretation already given may be summarized thus: Altho score is only a number indicating the credit value of an individual's performance, it results from the effectiveness of the individual's past environmental influences. That the effect which an environment can be expected to have varies with the nature of the environment and the length of exposure to it seems to be borne out by the immense amount of data already collected on the thousands of children tested by various mental tests. Without exception the average score for each age increases with increasing age. Here age is the measure of the exposure time. This does not hold for adults, because with adulthood comes specialization which means the restricting of environmental influences to a specific type. Increase could not, of course, continue indefinitely because there are degrees of perfection which cannot be exceeded.

Knowledge of a child's present and past life is of value as an aid in discovering which way he is headed, how fast he is traveling in that direction, and what his probable highest performance level will be eventually, providing he is stimulated by favorable environment. His direction, rate of travel, and probable limit can only be estimated in terms of his present performances taken as an index of the effectiveness of his past environmental influences. Barring abnormal circumstances, it would seem reasonable to assume that the rank order arrangement of performance levels of children of the same age would remain practically constant thruout life. Thus if in a given community the performance levels were determined for all children on their sixth birthday and they were listed in rank order, we would expect the rank order of the new performance levels of these children to be practically the same on each succeeding birthday. For convenience the performances of these children could be thrown into 100 different rank order groups. In the first or lowest group could be put that 1 per cent of the children who earned the fewest credit points, in the second group could be put the next lowest 1 per cent of the children, and in this way the groups could be constructed, each containing 1 per cent of the total group, until all were included. As each of these one hundred groups contains 1 per cent of the entire group, each may be referred to as a percentile. The particular position which each occupies may be indicated by numbering them in order from the bottom or lowest group up. The first percentile contains the children whose performance levels are so low that 99 per cent of all children the same age do better. The second percentile contains the children whose performance levels occupy such a position in the rank order arrangement that 1 per cent do worse and 98 per cent do better. In this way the number of each percentile group indicates exactly what per cent of children do worse and what per cent do better.

One value of this grouping is that it locates each child with respect to the entire group and permits a direct comparison to be made between any two children within the group. An individual's percentile is not

only a convenient way of presenting facts, but a precise statement of the exact location of his work compared with that of others exposed to the same environment for the same length of time. According to the theory of probability it is not likely that the increased exposure to environmental influences that is bound to come with increasing age will seriously alter the percentile rank order arrangement of the children. It should be expected that the performance level of all the children would be raised with increasing age, but that the rank order arrangement of these levels, as expressed in percentiles, would remain practically the same. Important changes in the environmental setting of a child or a group of children would undoubtedly produce a change in their performances and cause their rank orders to be shifted accordingly. The great variety and wide range of performances coming between the highest and lowest performance levels of children of each age produces considerable overlapping of the performance levels of children of different ages. This occurs because a given performance level is attained by some children in the course of, say, six years, while others do not attain this performance level until they have been exposed to the environment for seven years. There are others who need eight years of exposure. Certain individuals earn this performance level only after many more years, and some never earn it.

If the percentile ranks of children remain practically unchanged with increasing age, then the performance level of each child must increase by a relatively constant ratio varying directly with the increased length of time he is exposed to the environment. This then is the principle of human development which should serve as a guide in interpreting and using test results. If children of different ages make the same score on a test, this demonstrates them to have the same performance level on that test at the time the test is given, but of itself tells nothing about their probable future performance level.

Scientific classification of children requires that each group be standardized with respect to three criteria: first, that they all have the same performance level at the start; second, that they all be exposed to the same environmental, in this case educational, influences; and third, that the performance levels of all increase at the same rate so that their performance levels remain equal thruout and all finish at the same time with the same performance level. Mental and educational tests are the most reliable means available for determining a child's present performance level. Prediction for the rate of increase of his future performance level can be made only on the basis of measurable factors which have produced his present performance level. These measurable factors are the nature of his past environment and the length of time he has been exposed to it, usually indicated by his age.

A suddenly accelerated or retarded development is hardly probable. The development most likely to occur is merely a continuation of that which has characterized an individual's past, both as to its general nature, and as to its relative rate of increase with increasing age. It should not therefore be expected that two children of different ages having the same performance level would continue to have the same

performance levels, or would ever again have the same performance levels, any more than that it should be expected that two trains, one a freight and the other an express, going the same direction on parallel tracks would continue to be equidistant from the starting-point, or would ever again be the same distance from the starting-point before they reach their destination. At just what distance from the starting-point the express will pass the freight depends upon the rate of travel for both and how many hours the freight left the starting-point ahead of the express. If the express left first they never would be together. In this way each child may be thought of as either an express or a freight, depending upon the rate of progress of the other individual or individuals of the group with which he is being compared. Each child in the course of his life will have the experience of being both a freight and an express. There are enough children in the world that at every stage of each child's development his performance level is reached by children who left the starting-point after he did, some of them a long time after, and there are enough children traveling slower than he that at every stage of his development his performance level will reach that of children born before he was, some of them a long time before. It is therefore exceedingly important to recognize the fact that those children who make the same score and consequently have the same performance level on a test were caught just at the time they happened to be even, because just at that time they were passing one another in the various stages of their development. How fast they are traveling or what their rate of development is can be determined only by reference to the length of time consumed in arriving at this particular performance level.

Those children of this group who are of the same age are more likely to be traveling at the same rate and to have the same rate of development than any others. This gives a basis for selecting children to be put together in classes. They should be the same age and should have the same performance level. The performance level indicates the point at which to begin work in setting the educational environment for future development. The rate at which sections may be expected to develop within this environment varies inversely with the age of the children having the same performance level.

That this is exactly what happens is clearly demonstrated in the age-grade-score distribution of Table I. This table is compiled from data collected by the speaker in a survey of the city schools of Youngstown, Ohio. The National Intelligence Tests were used in testing all children from low third thru the eighth grade of twenty-five buildings. Age eight includes all children who had reached their eighth birthday, but who had not yet reached their ninth birthday on the day of being tested. All other age groups include children whose ages are similarly determined.

TABLE I. NATIONAL INTELLIGENCE TESTS, SCALE A: Average Score for each age-grade group having at least 20 children (8,167 children)

AGE	3B	3A	4B	4A	5B	5A	6B	6A	7B	7A	8B	8A	Average
7.....	37												40
8.....	36	49	65										44
9.....	30	46	62	79	98								55
10.....	27	41	56	72	87	99	104						66
11.....	25	38	51	67	81	93	105	115	136				83
12.....		34	46	59	76	89	95	107	118	132	142		94
13.....			44	57	69	80	93	104	112	126	136	142	106
14.....					65	77	87	99	105	111	127	137	111
15.....							86	87	101	112	122	127	109
16.....												134	114
17.....													114
Average....	31	44	53	69	80	89	97	105	111	123	130	136	Average

TABLE II. NATIONAL INTELLIGENCE TESTS, SCALE A: Percentage value of scores for each age-grade group compared to age average

AGE	3B	3A	4B	4A	5B	5A	6B	6A	7B	7A	8B	8A
7.....	93											
8.....	80	111	150									
9.....	55	84	113	144	178							
10.....	41	62	85	109	132	150	158					
11.....	30	46	61	81	98	112	126	139	157			
12.....		36	49	62	81	95	101	114	126	140	151	
13.....			42	54	65	76	88	98	106	123	130	134
14.....					59	70	80	90	106	100	114	123
15.....							80	80	93	103	112	117
16.....												118

Table I gives the average score for each age-grade group, having at least 20 children. Averages are not included where there are less than 20 children in a group. In 3B for age 7, the 37 means that is the average score for at least 20 children seven years old found in 3B. For age eight the 36 means the average score for at least 20 children eight years old found in 3B. The rest of the table is read in the same way. The last column at the right-hand edge of the table gives the average score of all children for each age of the entire city. The average for seven-year-old children is 40, for eight-year-old children is 44, etc., for each age down the column.

The numbers across the bottom of the table are the average scores for each grade. One interesting thing revealed by this table is that within each grade the average score decreases regularly with increasing age, i.e. there is an inverse ratio between score and age. That holds true with but one exception up to age 15. This constant decrease in score with increasing age within every grade is suggestive of the work-

ing of some definite principle which cannot safely be disregarded in efforts to group children scientifically.

If the numbers are read across the table, starting with 3B of any age thru the different grades, the scores increase regularly to the last score. This, of course, is the working of the same principle just referred to.

Table II is constructed from Table I by dividing the average score for each age into the average score of each grade group within that age, so that the 93 for seven-year-old children in 3B means that they did 93 per cent of the city average for seven-year-old children. The eleven-year-old children in 3B made an average score of only 30 per cent of the average score for all eleven-year-old children. The percentage decreases consistently with increasing age within any given grade up to age fifteen.

The averages for years fourteen and fifteen are not fully comparable with those for the other ages because only elementary school children are included. Many children fourteen and fifteen years of age were in high school. Thus the brightest fourteen and fifteen-year-old children are not included as probably only the duller ones had remained in the elementary schools.

This inverse ratio between age and performance level within any given grade must be reckoned with in the grading of children. It has

TABLE III.

Percentile Ranks of Scores on Indiana University Mental Survey Tests,
Schedule E, Cross-Out Test. Jan. 1, 1923.

SCORE	AGES					SCORE	AGES				
	9	10	11	12	13		9	10	11	12	13
0						51	85	77	62	52	39
1						52	87	80	65	56	42
2						53	88	83	67	58	44
3						54	89	85	70	60	47
4		1		3		55	90	86	73	63	50
5		1		3		56	90	87	75	66	52
6		1		3		57	91	89	77	69	56
7		1		5		58	93	91	80	72	58
8	2	1	1	5		59	94	92	82	74	60
9	5	1	1	5		60	94	93	84	76	62
10	7	1	3	7		61	95	95	85	77	66
11	7	1	5	7		62	95	95	87	80	69
12	1	1	8	8		63	95	95	88	82	72
13	1	3	8	8		64	95	95	89	84	74
14	1	5	8	8		65	96	96	91	86	77
15	1	6	8	8		66	97	97	92	88	79
16	2	6	1	8		67	97	97	93	88	81
17	3	1	1	8		68	97	97	94	89	82
18	4	1	1	1	2	69	98	98	95	91	84
19	4	1	1	1	2	70	98	99	96	92	85
20	5	2	1	1	2	71	98	99	96	93	86
21	6	3	1	1	4	72	98	99	97	94	89
22	8	4	1	1	6	73	98	99	98	95	92
23	10	5	3	1	8	74	99	99	98	96	94
24	12	6	4	2	1	75	99	99	98	97	95
25	15	7	4	2	1	76	99	99	98	97	96
26	17	8	4	3	2	77	99	99	98	97	97
27	19	10	5	4	2	78	99	100	99	98	97
28	21	12	6	4	2	79	99		99	98	98
29	23	13	7	5	3	80	100		99	99	99
30	25	15	9	6	3	81			99	99	99
31	30	16	11	6	4	82			99	99	99
32	33	19	12	7	4	83			99	99	99
33	37	20	13	8	5	84			99	99	100
34	40	23	16	10	6	85			99	99	
35	44	26	17	12	8	86			99	99	
36	48	30	19	14	9	87		100	99		
37	52	35	20	16	10	88			99		
38	55	38	24	18	12	89			99		
39	58	41	26	20	13	90			99		
40	61	44	29	22	14	91			99		
41	64	48	32	25	16	92			99		
42	67	51	35	27	18	93			99		
43	71	54	38	31	20	94			99		
44	73	57	42	33	23	95			99		
45	76	60	45	35	25	96			100		
46	78	64	48	38	28	97					
47	80	67	52	40	30	98					
48	81	70	54	43	32	99					
49	82	72	56	46	35	100					
50	84	74	59	48	37						
						Total number of cases	406	612	601	595	514

been a common practice to examine all the children of a given grade and then group them into classes on the basis of uninterpreted raw scores. In the light of these data, it would seem that such a procedure of grouping may nearly if not entirely defeat its own purpose.

Such a situation could easily occur and in fact does occur whenever sections of a given grade of a large school are formed by arranging the children in the order of the ranks of their scores and then assigning them to sections on the basis of these rank orders. The inevitableness of this can be shown best by analysis of some actual conditions. Last November thru the courtesy of Mr. R. N. Tirey, superintendent of the Bloomington schools, and his co-workers, and with the assistance of my class in Mental Measurements, all the children of the Bloomington public schools from the first grade up to and including the freshmen of the high school were tested. From the third grade up the children were tested by use of the Indiana University Mental Survey Test, Schedule

TABLE IV.—Age and Percentile Ranks of 30 Children from Grade 5B, who made the Same Score on a Mental Test.

AGE	Number of Children	Percentile Rank
8	1	2
9	1	67
10	12	51
11	7	35
12	2	27
13	2	18
14	2	?

E, the Cross-Out Test. The results obtained in this survey were combined with those reported from other cities, and a percentile table was constructed for each age from nine to thirteen inclusive. These were the only ages within which percentiles could be determined with a fair degree of reliability. The highest possible score on the Cross-Out Test is 100. The percentile ranks for this test are given in Table III. The scores are placed along the left edge of the table outside the vertical line. The numbers in the body of the table indicate the percentile rank of each score for each age. Under age nine is placed the percentile rank of each score made by nine-year-old children. The percentile ranks for the other years are placed in their respective columns. A score of 25 has a percentile rank of 15 when earned by a nine-year-old child, a percentile rank of 7 when earned by a ten-year-old child, of 4 for an eleven-year-old, of 2 for a twelve-year-old, and of only 1 for a thirteen-year-old child. The percentile rank of 15 assigned to the score 25 for a nine-year-old means that 14 per cent of nine-year-old children make lower scores, and that 85 per cent of nine-year-old children make higher scores. When a thirteen-year-old child makes a score of 25, his per-

centile rank of 1 means that less than 1 per cent of thirteen-year-old children make lower scores and that 99 per cent make higher scores.

In the Bloomington survey last November, 30 children from grade 5B made scores between 40 and 44. As 42 is the average score of this interval and all of the 30 children lie within two score points of it, it will simplify this discussion to assume that all 30 made the same score of 42. The distribution of these children according to age and percentile ranks is given in Table IV. The percentile ranks for the children eight and fourteen years old could not be determined, but that the eight-year-old would have a high rank and the fourteen-year-old a low rank cannot be doubted. It is not at all likely that these 30 children put into one class would constitute a very satisfactory class. They were caught by the test just at the time their performance levels happened to be the same. This will never happen again for these children. Yet they are the children that would be placed together under one teacher if the classes were selected on the basis of the rank order arrangement of scores of all children within the grade. A group of 30 children whose names had been drawn from a hat containing a shuffled list of all the children of the 5B grade would probably constitute about as satisfactory a class.

TABLE V.—Percentile Limits of Each Age Found in the Various Grades.
Bloomington Survey, Nov., 1922

Grades	6	7	8	9	10	11	12	13
8A						95 100	85 100	40- 100
8B						95 100	75 100	30 100
7A							60 100	20 100
7B					90 100	70 100	30 100	20 94
6A					70 100	60 100	20 100	5- 69
6B				95 100	45 100	20 100	5 94	5 59
5A				80 100	45 100	20 89	0 64	0- 54
5B				40 100	15 100	5 74	0- 49	0- 39
4A				25 94	5 89	5- 59	0- 44	0- 14
4B				5 94	0 84	0 29	0- 4	0- 19
3A			30 100	0 89	0 44	0 24	0 14	
3B	90 94	60 100	10- 100	0 59	0 39	0- 14	0 4	0- 4
2A		45 100	10 100					
2B	60- 100	10- 100	0 94					
1A	40- 89	5- 89	0 64					
1B	0 100	0- 94	0 49					

Table VI. Number of Cases Included in Table V.

GRADE	6	7	8	9	10	11	12	13
8A.....						1	6	33
8B.....						1	15	35
7A.....							23	27
7B.....					4	15	47	25
6A.....					4	25	28	22
6B.....				2	18	47	27	19
5A.....				1	21	32	16	13
5B.....				17	55	26	17	12
4A.....				30	29	25	13	4
4B.....				59	33	18	2	7
3A.....			25	35	21	15	5	
3B.....	1	6	87	47	18	2	1	1
2A.....		27	36					
2B.....	7	87	42					
1A.....	15	43	9					
1B.....	150	43	6					

After the Bloomington school children had been tested, the percentile rank was computed for each child and distribution tables were made for the city. Table V gives the lowest and highest percentile rank for the children of each age found in each grade. The numbers across the top of the table indicate the ages, those along the left side the grades, and those within the body of the table the percentiles. The percentiles go in pairs separated by a hyphen. The first one of each pair is the lowest percentile rank of any child within the age-grade group at the intersection of which it is located. The second of each pair is the highest percentile rank of any child within that age-grade group. Writing the minimum percentile in the upper left corner, and the maximum percentile in the lower right corner of the rectangular intersections of the age columns and grade rows, facilitates reading of either the minimum or maximum percentiles as they vary from grade to grade within each age column, or as they vary from age to age within each grade row. There were a few children under nine years of age in grades above the third, and there were also a number of children over eight years of age in grades below the third, but percentiles could not be determined for them because the mistake was made of testing the children by grades instead of by ages. One of the most significant features of this table is the consistent increase in size of percentile rank from the lowest to the highest grade within each age. In this connection it should be remembered that school grades are merely standardized educational performance levels. Some of the interesting facts revealed are such as these: that all eleven-year-old children with percentile ranks of 90 or more are found above the fifth grade; that no eleven-year-old child with a percentile lower than 20 is found above the low fifth grade; that whereas there is no nine-year-old child below the fortieth percentile in

the fifth grade, there is no thirteen-year-old child above the fortieth percentile in this same grade. Situations like the one mentioned last may be explained by application of principles of human development suggested earlier in this discussion, that the particular performance level an individual earns has of itself little value for purposes of prediction, and that the most important fact for purposes of prediction is the length of time a child is exposed to the environment before he reaches that performance level. Thus low percentile thirteen-year-old children and high percentile nine-year-old children are found functioning at the same educational performance level, because a given environment produces as much of an effect upon a high percentile child in nine years as it produces upon a low percentile child in thirteen years.

Table VI gives the number of children of each age-grade group reported in Table V.

SUMMARY

The detailed, rather theoretical discussion in the first part of this paper is the outgrowth of the author's efforts to interpret facts like those met with in the surveys reported in the latter part of the paper. This summary is a re-statement of the main points brought out by the discussion and the data of this paper.

1. For practical purposes it may be assumed that all children of a given community, and possibly of a given state, or even entire country, are subject to environments fundamentally so similar that they may be considered of equal value in stimulating development.

2. It may be assumed that when children of the same environment and of the same exposure time, i.e. same age, have markedly different performance levels, these differences constitute the best available index of their developability.

3. Score on a mental test is merely a numerical credit value of an individual's performance on that test at that time.

4. Score, i.e. performance level on a given test, has no predictive value except as interpreted by giving due consideration to the specific nature of the past environment to which the child has been subjected and to the length of time he has been exposed to that environment. If children have fundamentally the same environment, age is all that need be considered.

5. A rank order arrangement of scores or performance levels of children of exactly the same age indicates the relative effectiveness of their past environmental influences and probably is the most reliable index available of the effect that their future environment is likely to have upon them.

6. The percentile arrangement of scores within each age provides a very satisfactory method of reliably locating the relative position of each child's performance level with respect to that of all children having the same environmental influences. It is more convenient and more useful than any other method of presenting these facts. It has the added advantage of preserving the facts at their face value. It is not a theoretical manipulation of data. It is just a rank order arrangement with 100 ranks.

7. Percentile ranks have their greatest value when determined from a very large group of children for each age. Mental test standards presented in the form of a percentile table, like Table III, enable every user of the test to compute, directly from this table, the standard percentile rank of the scores made by his children. This permits him to compare his children with children from other cities on a standard basis just as easily as he compares his children with each other.

8. Tests should be given to children by ages and not by grades. In no other way is it possible to secure reliable percentiles on the performance of all children.

9. Scientific classification of children demands that each class should be made up of children who are of exactly the same age and who have the same level of performance. All children within a given school grade presumably have the same educational performance level, but they do not all have the same developability. Developability cannot be measured; it can only be inferred. Mental tests are a convenient aid in establishing indices of developability. Developability cannot, however, be inferred with certainty from a child's performance on only one mental test. If results from only one mental test are available and it is desired to use these in reclassifying a school, the following is a suggested procedure:

- First.* Determine from a standard percentile table the percentile ranks of all children.
- Second.* Make a list of the children in each grade. With each name give the child's exact age in years and months, and his standard percentile rank.
- Third.* Put together in classes those children from each grade who are of the same age and who have the same percentile ranks. This is seldom possible. It should, however, serve as a guide and, as an ideal, should be approximated as nearly as conditions permit.

Improving Instruction Thru Educational Measurement

FRANK W. BALLOU, *Superintendent of Schools, Washington, D.C.*

(Illustrated with lantern slides.)

THE ultimate purpose of all educational measurement is to increase the effectiveness of the instruction which the child receives, and thereby to improve the educational results which the child possesses from that instruction. Every administrative agency, every special teacher or supervisor, as well as as every regular teacher, all educational equipment; in fact, everything pertaining to the school system, is fundamentally for the purpose of providing the instruction which the child needs to make him individually and socially efficient.

Improvement of instruction thru educational measurement necessitates the following of a systematic procedure consisting of several steps: first, finding out what educational results are being obtained; second, evaluating those results by applying to them objective standards of achievement which have been established; third, getting the results of this evaluation to those responsible for the results or their improvement; and fourth, getting officers and teachers to attempt improvement.

This address is based on the methods pursued, and the results secured in the giving of the Curtis Standard tests in Arithmetic in Boston beginning in 1912 and extending over a period of four or five years. It indicates, in my judgment, a procedure which may be followed anywhere, under proper leadership, and in many subjects other than arithmetic.

STANDARD TESTS USED

The Curtis Standard Tests were used, of which the following are illustrations:

<i>Addition</i>		
Form 1	Form 2	Form 3
927	127	339
379	375	276
756	935	977
837	333	861
924	325	757
110	911	698
854	554	269
965	167	136
344	554	322

Subtraction

107795491
77197029

114957187
90271798

102142649
70428369

Multiplication

8246
29

8259
28

9237
27

Division

49)28420

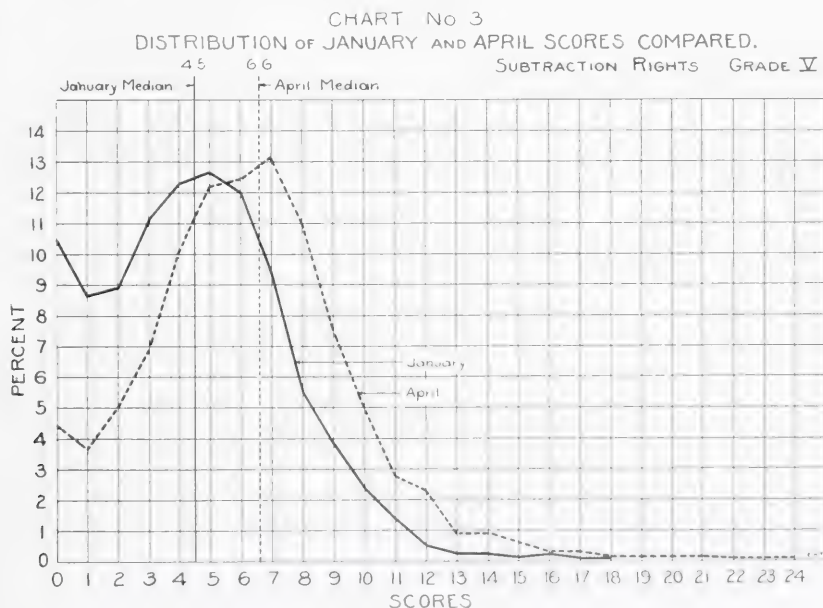
24)6984

26)7306

MEDIAN IS THE MEASURE

"Median" is the statistical term used to define the measurement of the central tendency of a group. "Median point" means the middle point in a statistical distribution, on either side of which there is an equal number of cases. The median achievement of a class means the achievement of that boy or girl who stands in the middle of the class when the pupils of the class are arranged in a straight line according to the achievement of each individual.

The median is illustrated by the following:



Solid line traces the January distribution.

Dotted line traces the April distribution.

Solid vertical line shows where the January city-wide median falls

Dotted vertical line shows where the April city-wide median falls.

The figures along the base line indicate the number of examples right.

The figures at the left indicate the per cent of children making each score.

Two important factors are illustrated by this chart; namely, the more extended distribution of the April scores and the advancement of the April median. For January the number of examples correct ranges from no examples to 18 examples, while for April the range is from no examples to 24 examples. This extension of the distribution is a mark of increased ineffectiveness in the teaching. The January median is 4.5 examples, the April median is 6.6 examples. This advancement of the median is a mark of improvement.

METHODS EMPLOYED TO IMPROVE RESULTS

Formal reports were made: (a) to teachers, (b) to principals, (c) to assistant superintendents and superintendent. Printed bulletins were distributed, objective standards were established, and systematic practice material was introduced. This address will deal with as many phases of these topics as time will permit.

OBJECTIVE STANDARDS OF ACHIEVEMENT IN ARITHMETIC

Addition				Subtraction			
Grade	Number of Examples Attempted	Per Cent of Accuracy	Efficiency	Number of Examples Attempted	Per Cent of Accuracy	Efficiency	
VIII....	12	80	33.1	12	90	33.9	
VII	11	80	30.1	11	90	28.5	
VI	10	70	12.5	10	90	33.8	
V	9	70	31.3	9	80	34.4	
IV	8	70	27.5	7	80	33.7	
Multiplication				Division			
Grade	Number of Examples Attempted	Per Cent of Accuracy	Efficiency	Number of Examples Attempted	Per Cent of Accuracy	Efficiency	
VIII	11	80	32.4	11	90	39.7	
VII	10	80	31.5	10	90	32.1	
VI	9	80	27.5	8	80	43.5	
V	7	70	38.7	6	70	39.9	
IV	6	60	35.2	4	60	40.2	

OBJECTIVE STANDARDS OF ACHIEVEMENT ESTABLISHED

The standard for each grade is based on the median achievement of all the pupils in that grade in 1915. It furnishes a definite goal for each pupil of the grade and one which he can reasonably be expected to reach with the right kind of training. A teacher should see that her class is trained to reach the standard and little if any more. A class has a satisfactory record when its median achievement approximates the standard in number of examples attempted, in per cent of accuracy, and in efficiency (the per cent of pupils who reach or exceed the standard in both speed and accuracy).

For example: The median which is taken as the standard for Grade eight shows that one-half of the class attempted at least 12 examples and did 80 per cent of them correctly; and that 33 per cent of the class reached or exceeded this standard.

The achievement of a class in speed should reach, but not materially exceed, the established standard for that grade. Achievement in accuracy should at least reach the established standard and may approach as nearly 100 per cent as possible. The efficiency of the class should at least reach the established standard and may approach as nearly 50 per cent as possible.

REPORTS TO PRINCIPALS AND SUPERINTENDENTS

(a) *Statistical.* The kind of a statistical report which was sent to each principal and superintendent before the close of school in June showing the standing of each class in his school as a result of the May tests is shown in the following illustration:

GRADE	Room	TEST NO. 1, ADDITION				TEST NO. 2, SUBTRACTION				TEST NO. 3, MULTIPLICATION				TEST NO. 4, DIVISION			
		Median		Efficiency		Median		Efficiency		Median		Efficiency		Median		Efficiency	
		Speed or Attempts	Accuracy or Rights			Speed or Attempts	Accuracy or Rights			Speed or Attempts	Accuracy or Rights			Speed or Attempts	Accuracy or Rights		
VIII.....	101	11	73.6	25.6		12	76.7	12.8		10	70.0	7.7		8	70.0	7.7	
VIII.....	119	10	67.3	2.5		9	76.0	5.0		9	62.0	12.5		7	73.3	10.0	
VIII.....	118	14	70.0	11.1		12	85.0	31.1		9	76.0	4.4		9	85.5	11.1	
VIII.....	227	13	87.8	34.9		12	95.5	39.5		10	84.5	20.9		10	96.7	34.9	
VII.....	77	11	70.0	12.2		11	76.0	19.5		9	72.2	12.2		7	78.6	14.6	
VII.....	133	9	76.7	20.5		9	82.3	20.5		9	73.8	15.9		7	84.8	11.4	
VII.....	235	13	80.0	39.5		12	96.3	58.1		10	85.7	46.5		10	95.0	48.8	
VII.....	136	11	78.3	31.7		12	86.9	56.1		10	79.2	36.6		10	97.5	61.0	
VII.....	238	14	79.3	32.6		12	94.5	62.8		11	88.5	46.5		9	100.0	41.2	

Speed or attempts—Number of examples completed by pupils. Accuracy or per cent of rights—Per cent of examples correctly solved. Efficiency—Per cent of class who reach or exceed the established standard.

Certain facts should be noted. No teacher's name appears on the report.

Low scores may not mean poor teaching; perhaps it is a class of backward pupils.

These reports are statements of fact about which the mother and the teacher can have no difference of opinion.

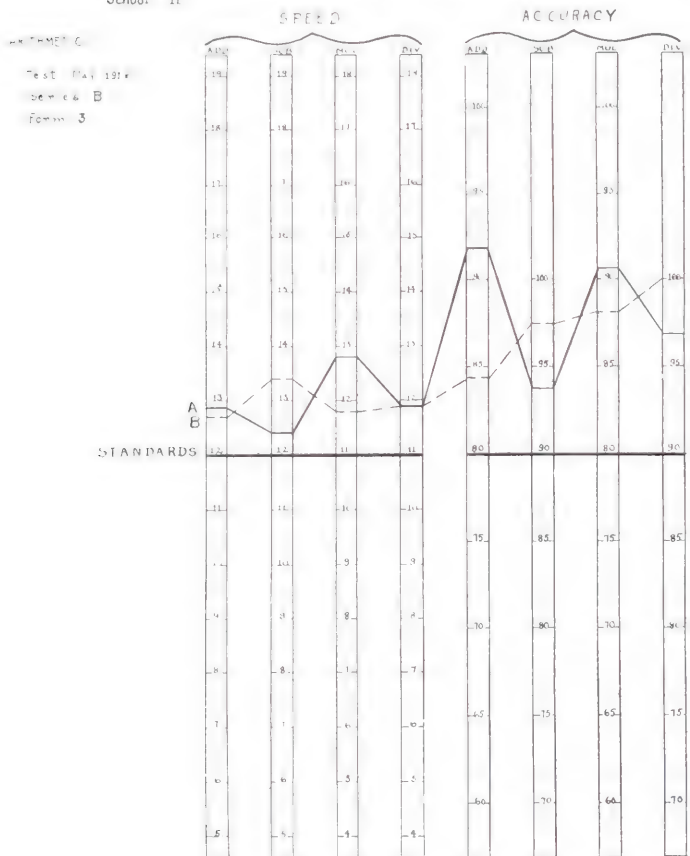
The principal can either justify the results or undertake to improve them if they cannot be satisfactorily explained.

(b) *Graphic.* The statistical results were also furnished the principals in graphic form, some of which will be here presented and explained.

A commendable school record:

DEPARTMENT OF EDUCATIONAL INVESTIGATION AND MEASUREMENT

School: II REPORT TO MASTER AND TO ASSISTANT SUPERINTENDENT Grade VIII



Teacher A's class is above the standard in speed in every process, altho the number of attempts never goes beyond 13. This class works with an accuracy of over 90 per cent in every process.

The record of Teacher B's class follows closely that of Teacher A's class. This class has a median achievement of 12.6 examples attempted in addition, 13.2 examples in subtraction, 11.7 examples in multiplication, and 11.8 examples in division. The accuracy also is above the standard in every instance.

This is a most commendable school record. It should be contrasted with that of School I. In School I one class conforms to the standard in speed and works with almost ideal accuracy, while the other class is far below the standard in both speed and accuracy. In School II, on the other hand, both classes conform to the standard in speed and work with almost ideal accuracy.

VARIATION IN CLASSES

DEPARTMENT OF EDUCATIONAL INVESTIGATION AND MEASUREMENT

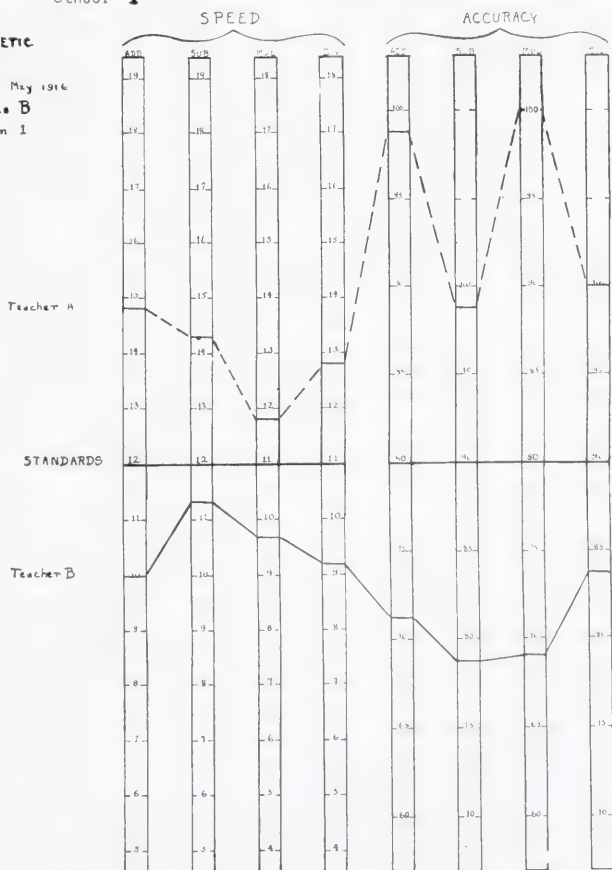
REPORT TO MASTER AND TO ASSISTANT SUPERINTENDENT

School I

Grade VIII

ARITHMETIC

Test May 1914
Series B
Form I



The median achievement of Teacher A's class is higher than the standard in both speed and accuracy in every process. It attempts 14.6 examples in addition, 14.4 examples in subtraction, 11.8 in multiplication, and 12.7 in division. Its per cent of accuracy is 98 in addition and subtraction, and 100 in multiplication and division.

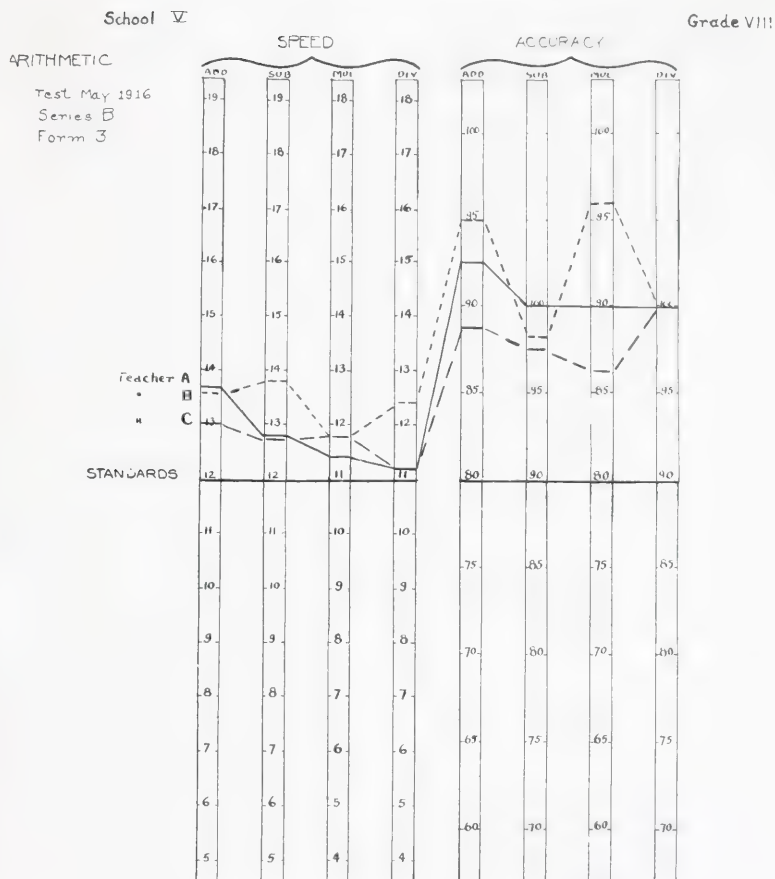
The median achievement of Teacher B's class is below the standard in both speed and accuracy in every process. It attempts 10 examples in addition, 11.3 in subtraction, 9.6 in multiplication, and 9.3 in division. Its accuracy ranges from 68 per cent in multiplication to 83 per cent in division.

This graph shows that there is a great difference in the arithmetical achievements of the two eighth-grade classes. The question for the principal is, "What is the reason for this in this particular class?" The results indicate unusual conditions and do not show whether these are due to poor teaching, with a failure to get at the individuals who are responsible for the low class median, or whether they are due to some legitimate cause.

SPLENDID SCHOOL RECORD

DEPARTMENT OF EDUCATIONAL INVESTIGATION AND MEASUREMENT

REPORT TO MASTER AND TO ASSISTANT SUPERINTENDENT



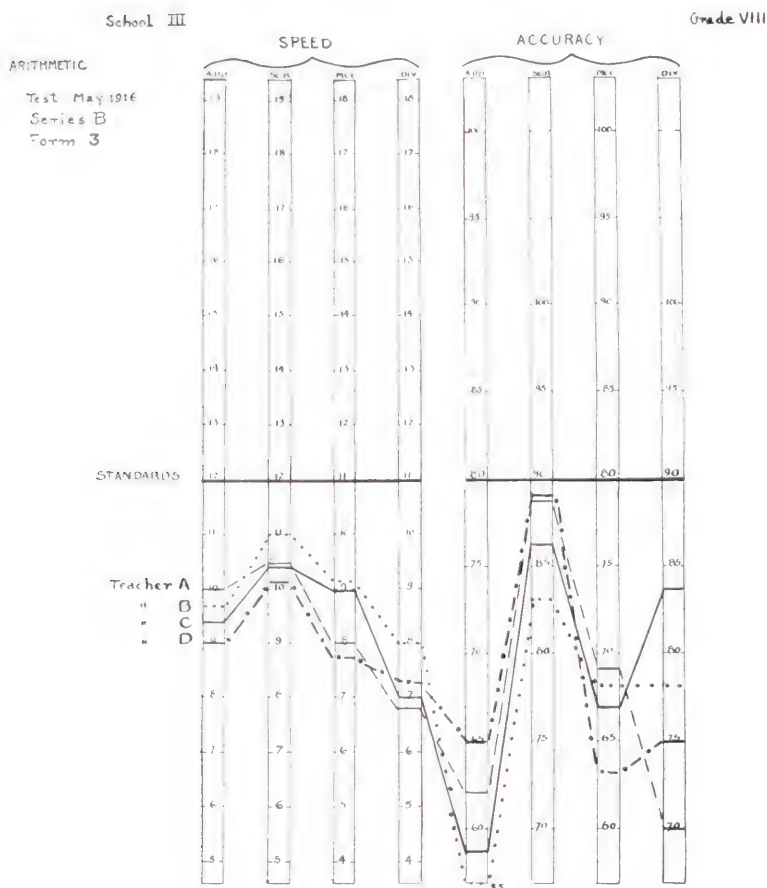
All three classes are above the standard in attempts in every process, but in no case does a class attempt more than 14 examples. Every class works more accurately than the standard requires. While the standard in accuracy in addition is 80 per cent, Class A achieves 92 per cent in that process; Class B, 95 per cent; and Class C, 89 per cent. The standard in subtraction is 90 per cent, while each of these classes works with an accuracy of 97 per cent or higher in this process. While the standard accuracy is 80 per cent in multiplication, 86 per cent is the lowest median achievement of these classes. All three classes work with a median accuracy of 100 per cent in division.

BULLETIN OF THE EXTENSION DIVISION

UNSATISFACTORY LOW SCHOOL RECORD

DEPARTMENT OF EDUCATIONAL INVESTIGATION AND MEASUREMENT

REPORT TO MASTER AND TO ASSISTANT SUPERINTENDENT



Teacher A's class attempts a median of only 10 examples in addition, 10.3 examples in subtraction, 8 examples in multiplication, and 6.8 examples in division. The accuracy of this class ranges from 62 per cent in addition to 89 per cent in subtraction.

Teacher B's class attempts a median of 9.6 in addition, 11 in subtraction, 9.1 in multiplication, and 8 in division, with an accuracy of 55 per cent in addition, 83 per cent in subtraction, 68 per cent in multiplication, and 78 per cent in division.

Teacher C's class attempts a median of 9.3 in addition, 10.3 in subtraction, 9 in multiplication, and 7 in division, with an accuracy of 59

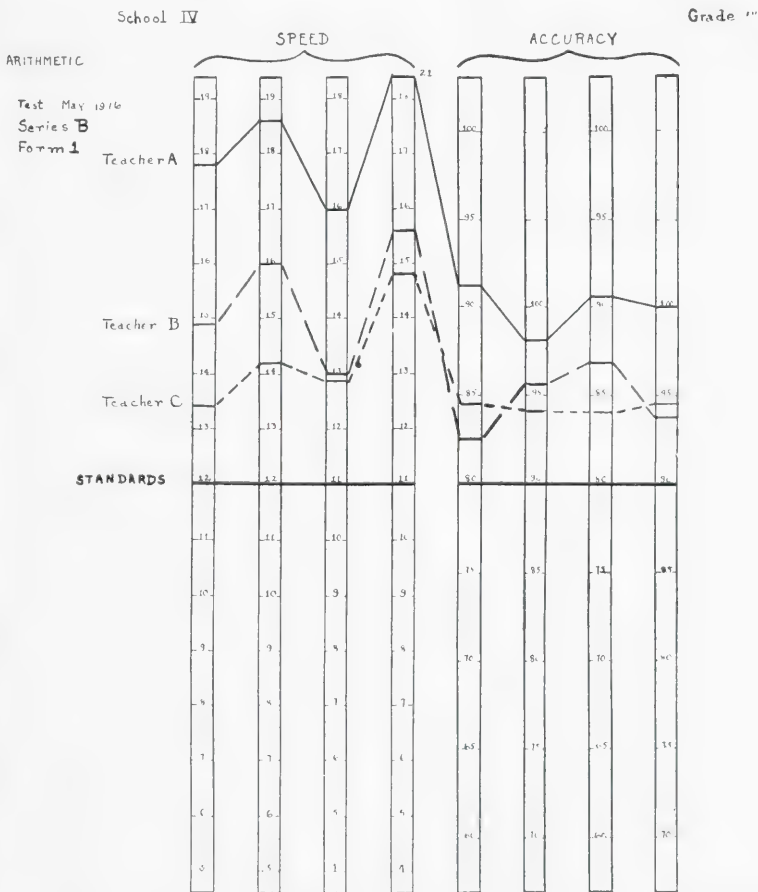
per cent in addition, 86 per cent in subtraction, 67 per cent in multiplication, and 84 per cent in division.

Teacher D's class attempts a median of 9 in addition, 10.1 in subtraction, 7.7 in multiplication, and 7.3 in division, with an accuracy ranging from 63 per cent in multiplication to 88 per cent in subtraction. All these class records are far below the standard for the grade in both speed and accuracy in every process.

ACCURACY SACRIFICED TO SPEED

DEPARTMENT OF EDUCATIONAL INVESTIGATION AND MEASUREMENT

REPORT TO MASTER AND TO ASSISTANT SUPERINTENDENT



Teacher A's class attempts 17.5 examples in addition, 18.5 examples in subtraction, 16 examples in multiplication, and 21 examples in division

IDEAL VS. OBJECTIVE STANDARD

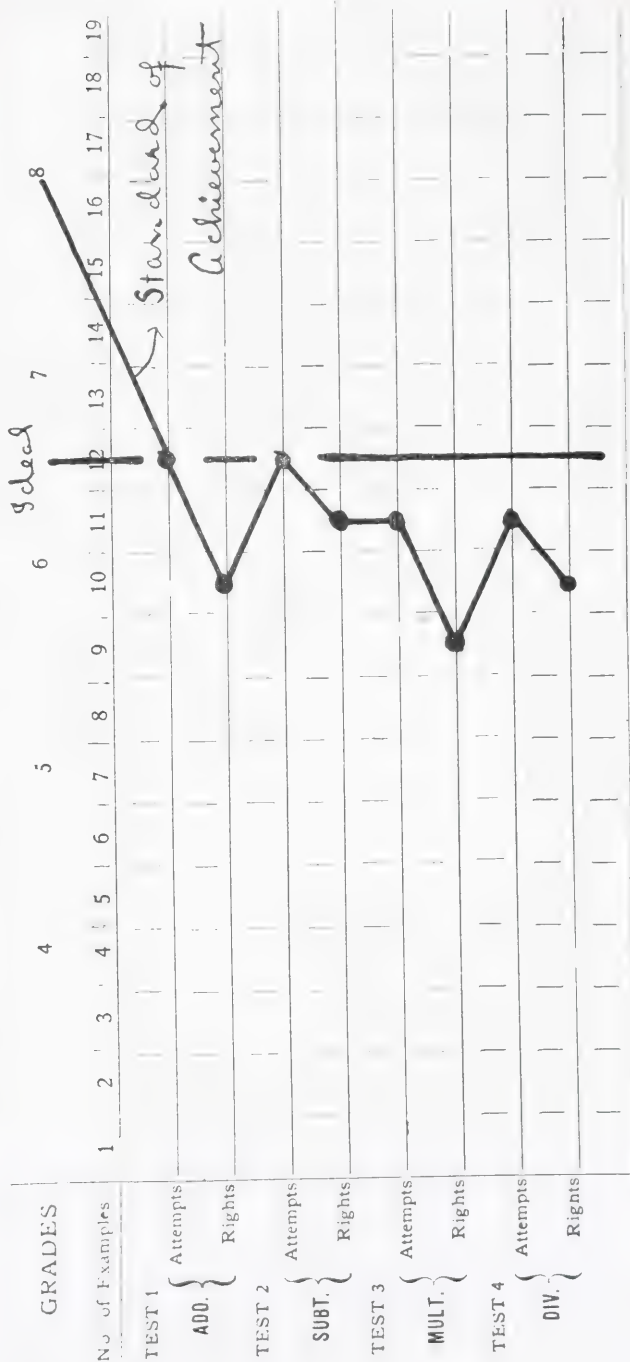
RECORD OF ARITHMETIC TESTS.

BOSTON
PUBLIC
SCHOOLSName *Ideal Present Standard*

Age at first trial

Boy or Girl

On last birthday



The card indicates to the teacher the pupil's ability or inability to perform the work in the four fundamental processes, so that she knows his needs and can govern her work accordingly.

The above graph shows the objective standard of achievement in the four fundamental processes in arithmetic for an elementary school graduate, based on what 50 per cent of the eighth-grade children in Boston have shown they are able to do. Therefore, it indicates a reasonable accomplishment, and one presumably within the grasp of a large proportion of eighth-grade pupils.

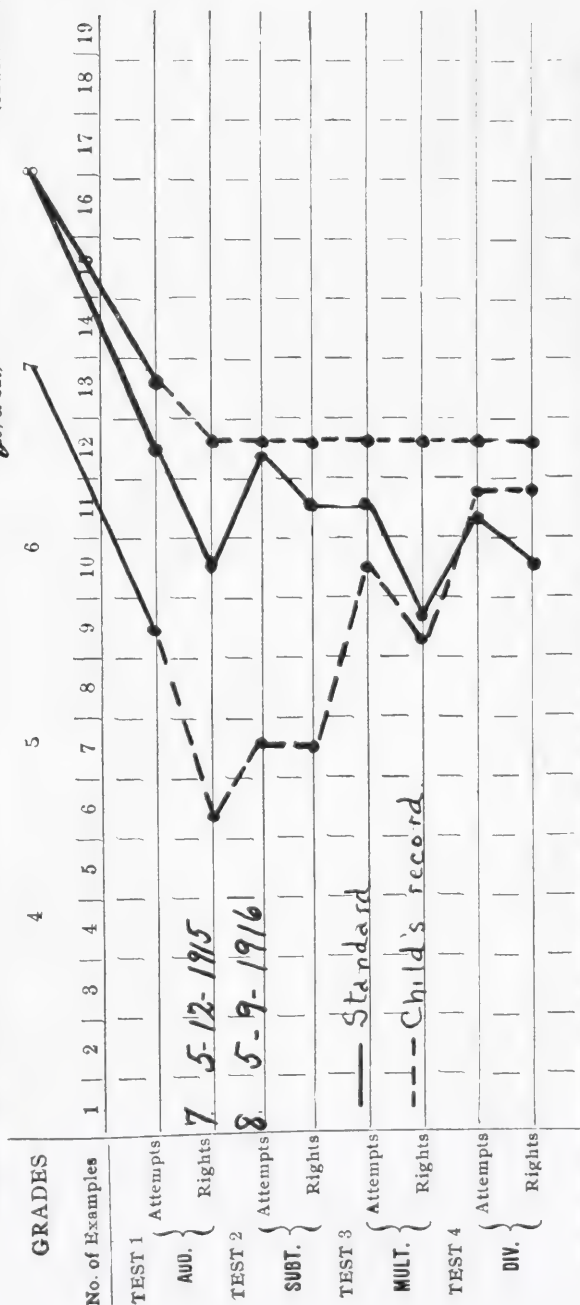
This standard is not the ideal, however. An ideal represents something striven for but not usually attained. In the case of the number of examples attempted the standard may well be considered the ideal. As to the degree of accuracy, altho a standard of 80 per cent has been set up as what may reasonably be expected, the ideal of accuracy is always to be 100 per cent. The graph shows the present standard in relation to the ideal. A record which is near the standard and which shows uniform ability as to the amount of work done, and also as to accuracy, approaches the ideal.

UNIFORM ABILITY IN OPERATIONS

It is desirable that a person should be equally proficient in the four fundamental processes of addition, subtraction, multiplication, and division. If he is, he will be able to attempt the same number of examples in each process in the Courtis Standard Tests and will work as accurately in one process as in another. As a matter of fact, however, too few pupils approximate to this achievement. The following picture shows the record of a child whose achievements approach the ideal in this respect.

RECORD OF ARITHMETIC TESTS

Name A. M. Age at first trial 13 (On last birthday)
 Boy or Girl Girl



INSTRUCTIONS On each horizontal line after the word Attempts, place a dot under the figure representing the number of examples you have tried, and on the line after the word Rights, place a dot under the figure representing the number of examples that you had right. Beginning with the figure at the top that represents your grade draw a line through all the dots that you have made.

When she first took the test she showed uneven ability in the processes. She was slow in addition as well as inaccurate. She was accurate in subtraction but slow. Her multiplication score was up to eighth-grade standard in the number of examples correct, while the division score was up to the eighth-grade standard in speed and showed perfect accuracy. In the second test she showed remarkable uniformity in the four processes, attempting 12 examples in every process but addition, in which she attempted 13, and in every case getting 12 examples correct.

The second score is an unusual one and shows an ideal toward which to work. While the number of examples attempted is at or near the standard, there is at the same time almost perfect accuracy. Just enough emphasis has been placed on the deficiencies in addition and subtraction to eradicate them and bring ability in those processes up to the ability in multiplication and division. The record of this girl is the kind of standard to which to look forward. It would constitute a better standard than the present one.

VARYING ABILITY IN OPERATIONS

In contrast to the foregoing, the following is an example of extreme variability in ability to add, subtract, multiply, and divide.

RECORD OF ARITHMETIC TESTS.

Name H

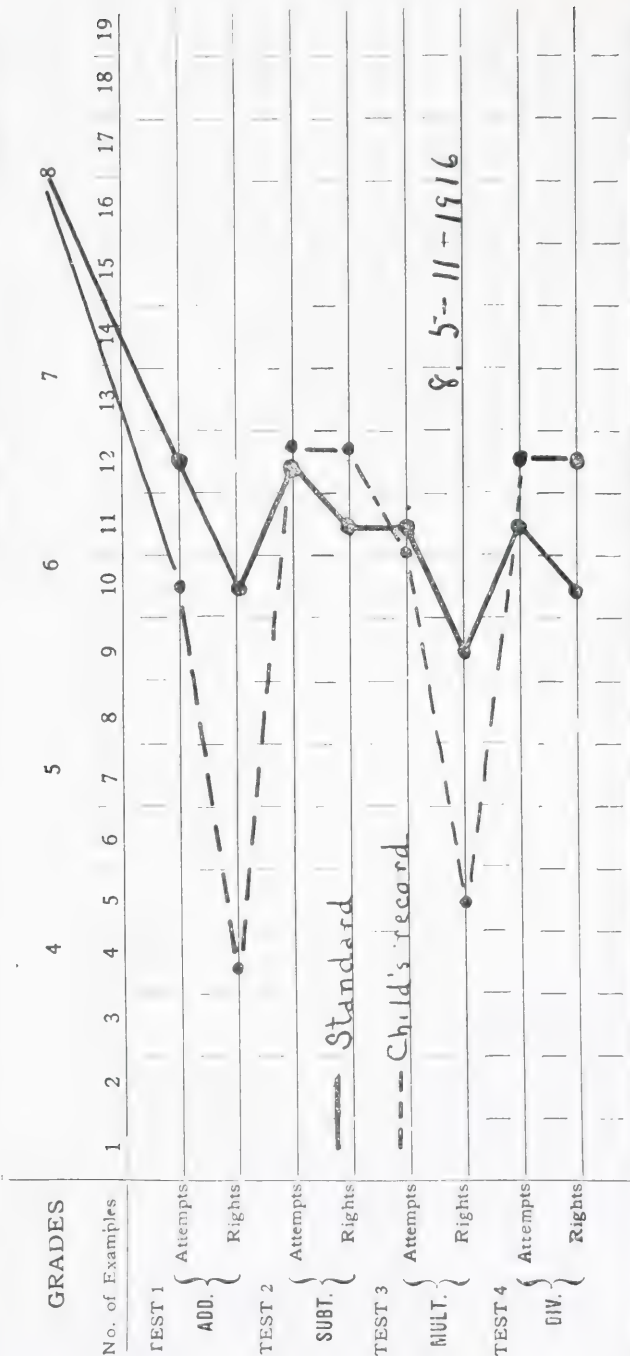
K

girl
(Boy or Girl)

Age at first trial

16

(On last birthday)



INSTRUCTIONS. On each horizontal line after the word Attempts, place a dot under the figure representing the number of examples you have tried, and on the line after the word Rights, place a dot under the figure representing the number of examples that you had right. Beginning with the top figure at the top that represents the grade draw a line through all the dots that you have made.

The record shows varying ability in the four fundamentals. The addition score is 10 examples attempted with only 4 right, a record below the eighth-grade standard both in speed and accuracy; the multiplication score is 11 examples attempted with only 5 right, a record up to the eighth-grade standard in speed but far below it in accuracy. The subtraction and division scores are above the eighth-grade standard. In both these processes the girl attempts 12 examples and performs the work with perfect accuracy.

This card shows clearly where drill is needed. If this child had taken the tests in earlier years, the weakness in addition and multiplication would have shown in the previous records, thus making it possible for the teacher to correct it. As it is, this girl graduates from an elementary school with no greater ability to do arithmetical problems than that represented in her addition and multiplication scores. Only individual drill on addition and multiplication will help this girl.

ERADICATION OF A DEFICIENCY

When a pupil finds difficulty in performing examples in any one of the four processes, he needs special help in order to get at the root of his trouble. The teacher must first find out what his trouble is and then apply the remedy. No amount of general class drill will eradicate his particular deficiency. Each pupil must be given the opportunity to practice on the process with which he finds most difficulty.

The following is a record of a pupil whose teachers helped him to overcome difficulties:

By examining the record of this boy it will be seen that there was constant improvement in both speed and accuracy in every test. In the first test the record shows particularly slow and inaccurate work in addition and division. In addition, 4 examples were attempted with only 2 right, while in division only 4 were attempted with 0 right. While the subtraction and multiplication scores were not so bad as these, they were far below the standards for the grade. In the second test the record shows improvement in every process. That in division was particularly good while that in addition was only slight, for the work in this process was still very inaccurate. In the third test perfect accuracy has been attained in every process, altho with the exception of division the speed is not quite up to the eighth-grade standard. The graph on the following page shows how the uneven line of achievement has been gradually straightened in the three successive tests.

This record proves the value of its existence. As a result of the first test, the individual's troubles in the fundamentals were diagnosed and treatment applied where it was most needed; i.e. in division. The second test showed the efficacy of this treatment. Addition, however, still made a poor showing. The third test showed that proper treatment had been applied, for this deficiency was overcome. The speed still was below the standard. However, perfect accuracy was reached and but little extra drill on addition and multiplication would be needed to make of this an ideal score.

PERSISTENCE OF A DEFICIENCY

In contrast to the foregoing, the following picture shows a record of a pupil whose teachers have failed to help her to overcome her difficulties:

RECORD OF ARITHMETIC TESTS.

Name R — girl Age at first trial 12
(On last birthday)
(Boy or Girl)

School District School City

DATE	GRADE	SERIES B									
		TEST 1 ADD.		TEST 2 SUBT.		TEST 3 MULT.		TEST 4 DIV.			
		Attempts	Rights	Attempts	Rights	Attempts	Rights	Attempts	Rights		
1914-10-14	7	9	5	11	10	10	8	11	7	1st Trial	
1915-5-13	7	15	8	16	15	13	8	14	11	2nd Trial	
1916-5-8	8	10	5	12	11	12	12	14	14	3rd Trial	
										4th Trial	
										5th Trial	
										6th Trial	
										7th Trial	

The girl's first test showed that at the beginning of the seventh year her subtraction and multiplication scores were already up to seventh-grade standards; her division score, altho up to seventh-grade standard in speed, was rather inaccurate; in addition the record showed that she was slower and more inaccurate than the average fifth-grade child. At the end of that year, her second test score showed that she had increased her speed in every process altho her accuracy had not improved; in fact, in multiplication she was more inaccurate than in the first test. At the end of the eighth year the third test showed work that was at or above the eighth-grade standard in all processes but addition. In this process the deficiency was still apparent.

The record showed both the seventh-grade teacher and the eighth-grade teacher that this child was meeting some difficulty in her addition work that prevented her doing accurate work. An effort should have been made to discover that difficulty and eradicate it.

STRIKING ACCURACY

Perfect accuracy is always the ideal to be striven for.

RECORD OF ARITHMETIC TESTS.

Name A-S- and O-P- Age at first trial On last birthday
 School District City Boston School City Boston

DATE	GRADE	1st Trial	2nd Trial	3rd Trial	4th Trial	5th Trial	6th Trial	7th Trial	SERIES B							
									TEST 1 ADD.		TEST 2 SUBT.		TEST 3 MULT.		TEST 4 DIV.	
									Attempts	Rights	Attempts	Rights	Attempts	Rights	Attempts	Rights
<u>A-S-</u>																
<u>1915-5-12</u>	<u>7</u>								<u>21</u>	<u>14</u>	<u>12</u>	<u>12</u>	<u>11</u>	<u>10</u>	<u>12</u>	<u>10</u>
<u>1916-5-9</u>	<u>8</u>								<u>16</u>	<u>16</u>	<u>14</u>	<u>14</u>	<u>14</u>	<u>13</u>	<u>13</u>	<u>13</u>
<u>O-P-</u>																
<u>1915-5-11</u>	<u>7</u>								<u>10</u>	<u>10</u>	<u>12</u>	<u>11</u>	<u>10</u>	<u>8</u>	<u>10</u>	<u>8</u>
<u>1916-5-9</u>	<u>8</u>								<u>11</u>	<u>11</u>	<u>12</u>	<u>12</u>	<u>12</u>	<u>12</u>	<u>12</u>	<u>12</u>

The seventh-grade record of A..... S..... in 1915 shows good accuracy except in addition. The speed is up to or above eighth-grade standard in every process. Her eighth-grade record in 1916 shows perfect accuracy. There has been a slight gain in attempts except in addition, where the accuracy is attained at the expense of speed, which in this case could afford to suffer. The record of O..... P..... in 1915 shows good accuracy with the attempts up to or above seventh-grade standard except in addition. In 1916 he attempts just 12 examples in every process except addition, where he attempts 11, and works with perfect accuracy in all four processes.

These records show striking accuracy. The May, 1916, record of O..... P..... is almost ideal, for altho the attempt score falls slightly below the standard in addition, the accuracy is perfect. Just enough drill work has been afforded to keep the attempts near the standard. The 1916 record of A..... S..... is ideal as regards accuracy. The change in the addition score from 21-14 to 16-16 is especially noteworthy. The attempts, however, exceed the standard, showing that the girl did not need so much drill as she was given in order to maintain her speed, which was already near the standard.

ANOTHER CASE

School

[illegible]

The first test showed that this boy when in grade six possessed ability above that of the majority of eighth-grade pupils in all processes except multiplication. He worked with almost perfect accuracy. This striking accuracy was repeated in both the second and third tests. There was little gain in attempts. In fact, there was a falling off in the number of examples attempted in division. At the end of the eighth-grade year he was above the eighth-grade standard in speed and worked with splendid accuracy.

This boy showed in his first test a reasonable achievement in the four processes; in fact, a higher achievement than 50 per cent of the eighth-grade pupils in the city attained. There was no necessity of his receiving more than a very small amount of drill on the fundamentals. Apparently he received a very slight amount, as his last test showed little gain in ability over the first test. This is as it should be. The energy necessary to raise this child's score higher was being rightfully applied to other more important things.

EXCEPTIONALLY HIGH AND LOW ABILITY

Examples of two kinds of records follow: those of pupils whose native ability is high and those of pupils whose native ability is low. Naturally children of the first type require much less training to bring them to a certain level than the children of the second type require. The former will profit far more by a class drill than the latter. In this difference in capacity lies the greatest reason for the need of training suited to the individual.

RECORD OF ARITHMETIC TESTS.

223
Name a. B.

R m
T s

Age at first trial Girls
(Boy or Girl)

{13
12

(On last birthday)

School District

School

City Boston

DATE	GRADE	SERIES B							
		TEST 1 ADD.		TEST 2 SUBT.		TEST 3 MULT.		TEST 4 DIV.	
		Attempts	Rights	Attempts	Rights	Attempts	Rights	Attempts	Rights
a. { Oct. 1914 May 1915	1st Trial	3	0	4	3	6	3	5	3
	2nd Trial	8	5	10	5	9	8	6	2
	3rd Trial								
B { Oct 1914 May 1915	4th Trial	13	12	19	18	10	9	11	11
	5th Trial	24	23	24	24	18	15	20	20
	6th Trial								
	7th Trial								

In the first test the girl designated as "A" had a low achievement in every process. In no process was her achievement so high as the fourth-grade standard. In the second test her record showed some increase in the number of attempts, but, on the whole, none in accuracy. In no process was the standard for grade seven reached. On the other hand, the girl "B" had a high achievement in the first test; in fact, in this test her achievement exceeded the eighth-grade standard in all processes but multiplication, which was already above seventh-grade standard. In the second test she performed from 5 to 11 more examples in each process than she did in the first test, with no loss in accuracy.

A, who needed much more drill work in this year than B, got in all probability the same amount. B, possessing a high achievement to begin with, and needing practically no further drill as it was evident that there were no serious difficulties for her to overcome, profited far more from the class drill than A possibly could until her own individual difficulties were overcome. Thus A did not get so much drill as she needed, while B was forced to spend, on drill which was unnecessary for her, time that could have been more profitably employed on some other work.

In this test all these pupils achieved a score far beyond the standard set for eighth-grade pupils. They succeeded in performing correctly most of the examples which they attempted.

Pupils who can achieve such high scores as those shown here demonstrate that their native ability is high. Hence they need a minimum of training in order to bring to a desirable height their ability in this particular subject. When that desired height has been reached (in this case the objective standards of achievement in the four fundamental processes) then time available for further training should be devoted to other branches of learning.

RECORD OF ARITHMETIC TESTS.

Name *Low Scores.*

Age at first trial
(On last birthday)

(Boy or Girl)

School District

School

City

Boston

SERIES B

DATE	GRADE		TEST 1 ADD.		TEST 2 SUBT.		TEST 3 MULT.		TEST 4 DIV.	
			Attempts	Rights	Attempts	Rights	Attempts	Rights	Attempts	Rights
<i>May 1916</i>		1st Trial								
		2nd Trial								
	<i>C-E-</i>	3rd Trial	<i>10</i>	<i>3</i>	<i>9</i>	<i>8</i>	<i>7</i>	<i>3</i>	<i>7</i>	<i>2</i>
	<i>A-D-</i>	4th Trial	<i>5</i>	<i>3</i>	<i>7</i>	<i>4</i>	<i>5</i>	<i>3</i>	<i>4</i>	<i>3</i>
	<i>V-A-</i>	5th Trial	<i>8</i>	<i>1</i>	<i>8</i>	<i>4</i>	<i>5</i>	<i>1</i>	<i>7</i>	<i>5</i>
	<i>E-M-</i>	6th Trial	<i>4</i>	<i>1</i>	<i>3</i>	<i>2</i>	<i>2</i>	<i>0</i>	<i>2</i>	<i>0</i>
		7th Trial								

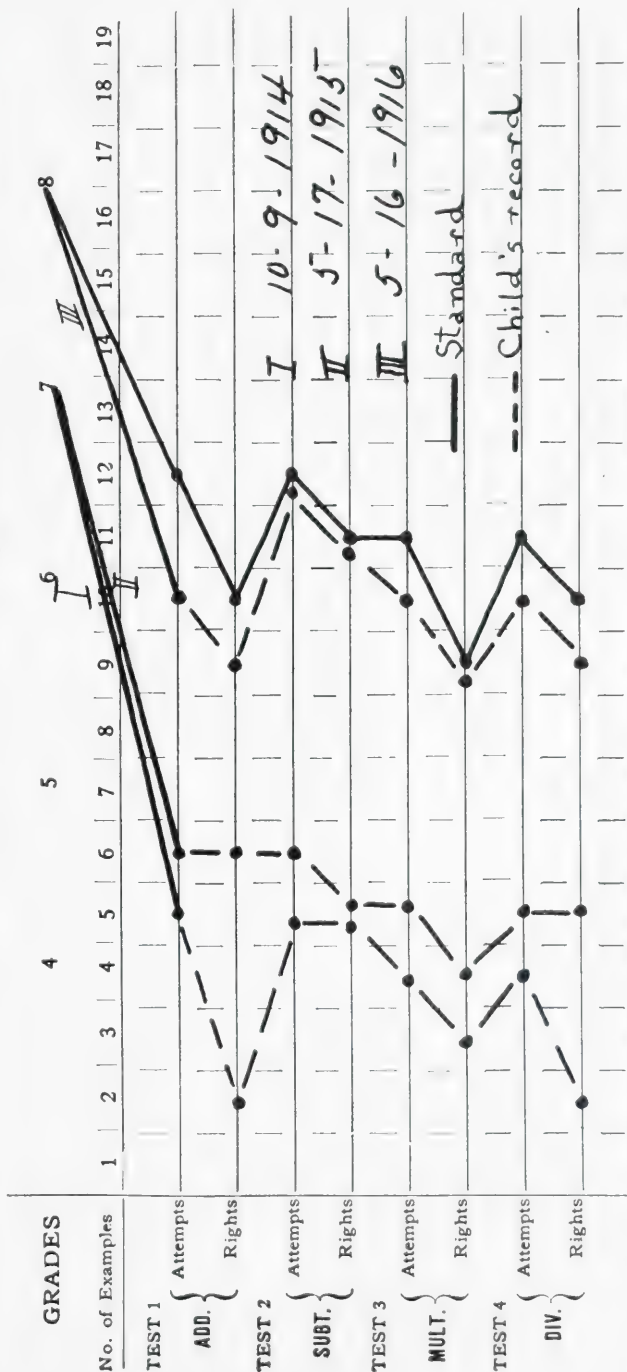
In this test C..... E..... did not reach the sixth-grade standard in any process; no one of the other pupils whose records are shown here succeeded in achieving in any process what 50 per cent of the children in grade four could do, with the one exception that V..... A..... could perform in division a little more than the fourth-grade median achievement.

Pupils who achieve such low scores probably demonstrate that their native ability is low. Hence they are individuals who certainly need a maximum of training in order to attain a satisfactory ability. If one contrasts these records with the foregoing group of high records, one can readily see the need of drill suited to the individual, as class drill in the fundamentals, cannot be adapted to the needs of all members of the class.

NOTEWORTHY ANNUAL GROWTH

A few graphs will show how pupils progress satisfactorily from year to year:

RECORD OF ARITHMETIC TESTS.

223
BOSTON
PUBLIC
SCHOOLSName MBgirl
(Boy or Girl)Age at first trial 11
(On last birthday)

INSTRUCTIONS. On each horizontal line after the word Attempts, place a dot under the figure representing the number of examples you have tried, and on the line after the word Rights, place a dot under the figure representing the number of examples that you had right. Beginning with the figure at the top that represents your grade draw a line through all the dots that you have made.

INTERPRETATION OF THE ABOVE RECORD

This graph shows the results in three successive tests of a girl who was eleven years old when she took the first test. The first test was in October, 1914, when she was beginning the work of grade seven. The second test was at the end of that year, May 17, 1915. The third test was in May, 1916, when she was about to graduate.

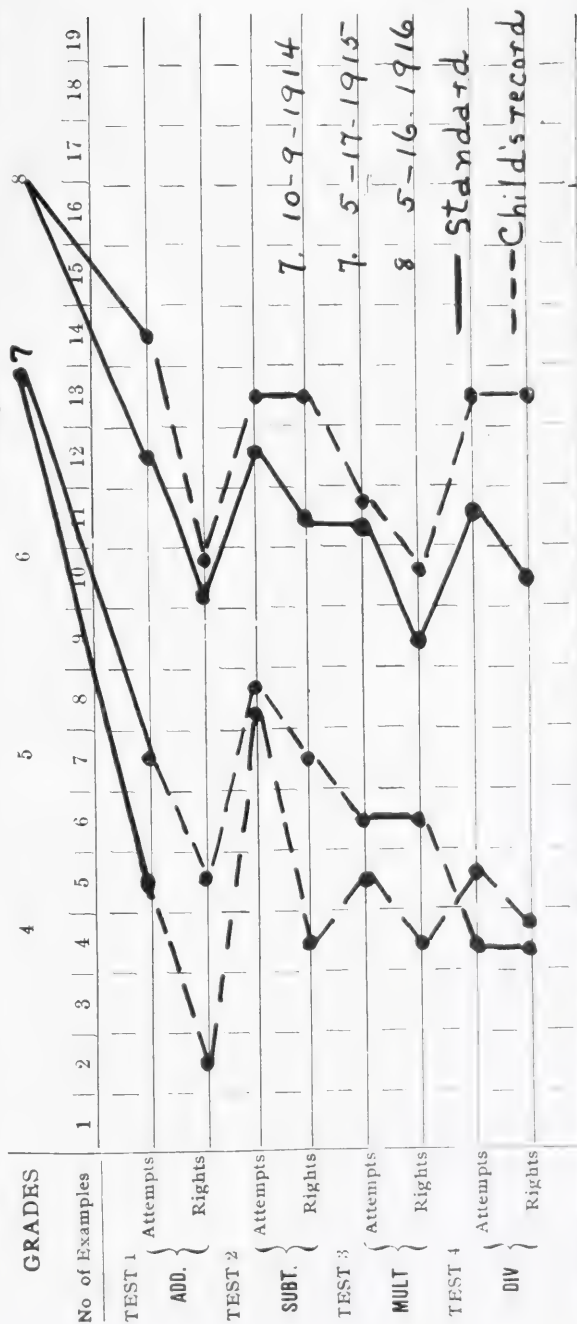
In the record of the first test the graph shows an uneven line due to the inaccuracy in addition and division. The accuracy in subtraction and multiplication is good, but the number of attempts in each process is only 4 or 5, which means that it never reaches the achievement of the ordinary fifth-grade child. In the second test an improvement in speed and accuracy is shown. The girl attempts 5 or 6 examples in each process and performs correctly all that she attempts. The speed is still below the fifth-grade standard. The improvement in addition and division, which were inaccurate before, is especially noticeable. In the third test there is a much larger number of examples attempted in each process. In this test the score conforms rather closely to the eighth-grade standard. The high quality of accuracy achieved in the second test has suffered somewhat.

The annual growths in achievement shown by this card are very commendable. In the seventh grade the accuracy is improved. There must have been special emphasis placed on addition and division to improve those scores so much. In the eighth grade the speed is improved. The growth attained in grade eight could hardly be better, for in this one year, from being able to do only work that would not compare favorably with the work of fifth-grade children, this girl increases her ability so that it almost reaches the standard set for eighth-grade children.

BOSTON
PUBLIC
SCHOOLS

RECORD OF ARITHMETIC TESTS

Name A. O. N., Boy Age at first trial 10
(On last birthday)



INSTRUCTIONS: On each horizontal line after the word Attempts, place a dot under the figure representing the number of examples you have tried, and on the line after the word Rights, place a dot under the figure representing the number of examples that you had right. Beginning with the figure at the top that represents your grade draw a line through all the dots that you have made.

INTERPRETATION OF THE ABOVE RECORD

This graph shows the records of three successive tests taken by A..... O'N....., the first and second in October, 1914, and May, 1915, when he was a pupil in grade seven, and the third in May, 1916, when he was a pupil in grade eight.

In the first test this boy attempts only a small number of examples in each process. Moreover, his work in addition and subtraction is decidedly inaccurate. In the second test he shows improvement in both speed and accuracy in every process except division where he attempts only four examples. He is still below the standard in speed for his grade. In the third test he has made such progress that his record is beyond but near the eighth-grade standard.

The growth from year to year in this child's ability is good. He entered grade seven with arithmetical ability lower than that of half the children in grade four. When he left grade eight, his ability had been increased so that it was higher than that of half the children in grade eight. It approximated the eighth-grade standard.

SYSTEMATIC PRACTICE MATERIAL INTRODUCED

Undoubtedly, the most important means of improving arithmetic instruction in the four fundamental operations has been the introduction of systematic practice material or exercises which take the place of the former general class drill in which every child practiced on the same exercise. The Curtis tests have demonstrated conclusively that the mass or class drill exercises in addition, subtraction, multiplication, and division are ineffective and uneconomical. Mass drills provide unnecessary practice for the more capable third of the class and hence are uneconomical; they do not provide enough practice for the less capable third of the class and hence are ineffective. An economic distribution of drill requires that it provide a means of reaching the individual needs of the 40 pupils of the class.

In an address before the Masters' Association on October 13, 1914, the writer defined the essential characteristics of satisfactory practice material in arithmetic.

1. It must be easily handled by pupils and teachers within a limited time allotment, say ten minutes.
2. It must provide for the correction of the problems by the pupils and not by the teacher.
3. It must provide for each child a kind of work suited to his particular needs.
4. It should provide the pupil with a method by which he can keep a record of his daily progress.
5. It should be progressively graded so that the child may pass from one kind of exercise to another as his ability increases, always finding the kind of exercise in which he needs practice.
6. It should provide a means of keeping a record of the progress of the class.

IMPROVEMENT IN SCHOOL SYSTEM AS A WHOLE

A splendid opportunity presented itself in 1915 to measure the results of the systematic work which had been carried on in educational measurement in Boston because the testing work had by that time become city-wide, and because the schools could be easily classified as to participation in the measurement work. Schools were grouped into Group A, Group B, and Group C.

Group A schools in the following discussion are those in which the tests have been given since 1912. In Group A there are 29 schools and 18,391 pupils represented.

Group B schools are those which were added during the second year or the first half of the third year of testing. They are schools which have been tested from one to two years, and are schools in which the effects of the testing may legitimately vary. In Group B there are 17 schools and 15,241 pupils represented.

Group C are those in which the tests were given for the first time in May, 1915. They are schools which have not been affected by the Curtis testing work in the city, except in so far as general discussion of the work has affected individual teachers or principals. It should be said, however, that principals of these schools have had opportunity to introduce the practice material, and several of them did so. If these schools had not been at all affected by the system of educational measurement in the city, the superiority of the schools tested would have been even greater than it is. In Group C there are 17 districts and 11,836 pupils represented.

The comparison of the results achieved in 1915 by these groups of schools is as follows:

Addition

Comparison of Median Scores of Group A, Group B, and Group C Schools,
May, 1915

GRADE	GROUP A		GROUP B		GROUP C	
	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy
VIII.	14.4	79.07%	13.1	77.84%	12.6	76.41%
VII	12.9	77.50%	11.6	74.15%	11.6	74.24%
VI	11.5	78.17%	10.4	72.72%	11.1	74.14%
V	9.3	72.93%	9.6	71.61%	9.4	68.64%
IV.....	8.2	70.00%	7.6	65.15%	7.7	63.11%

Subtraction

Comparison of Median Scores of Group A, Group B, and Group C Schools,
May, 1915

GRADE	GROUP A		GROUP B		GROUP C	
	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy
VIII	14.6	89.73%	13.1	88.77%	12.4	86.92%
VII	12.9	87.57%	11.7	86.26%	11.5	84.92%
VI	11.5	88.01%	10.7	85.69%	11.0	88.73%
V	9.3	85.14%	9.4	84.44%	9.00	80.64%
IV	7.9	83.63%	7.5	81.56%	7.1	78.28%

Multiplication

Comparison of Median Scores of Group A, Group B, and Group C Schools,
May, 1915

GRADE	GROUP A		GROUP B		GROUP C	
	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy
VIII	12.1	82.58%	11.2	81.49%	10.8	80.42%
VII	10.8	81.58%	10.3	79.28%	10.1	80.08%
VI	9.5	78.45%	8.8	76.67%	9.4	79.93%
V	7.6	75.69%	7.7	76.01%	7.3	70.97%
IV	6.1	70.95%	5.9	67.45%	5.8	63.61%

Division

Comparison of Median Scores of Group A, Group B, and Group C Schools,
May, 1915

GRADE	GROUP A		GROUP B		GROUP C	
	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy
VIII	12.7	93.46%	12.0	92.08%	11.0	89.72%
VII	10.9	89.34%	9.6	87.63%	9.3	86.37%
VI	8.9	88.09%	7.9	85.02%	8.6	86.72%
V	6.2	78.98%	6.7	80.05%	6.2	73.68%
IV	4.7	65.85%	4.6	59.11%	4.8	63.11%

Undoubtedly, the objective standards of achievement, the various reports to teachers, to principals, and to the superintendent, the printing of bulletins for general distribution, and the introduction of scientific practice material, have all contributed something toward the improvement of results in arithmetic. Of more pronounced effect probably than any of these factors, however, has been the stimulation among teachers of an inquiring attitude toward the whole problem of arithmetic instruction. The results from the tests have shown that the problem of arithmetic teaching is not yet solved; and they have prompted many teachers to study their own work as the first step toward improving methods of instruction.

These results seem to prove two important propositions: (1) That the scientific measurement of educational results is possible and practicable; and (2) that educational measurement may be a means of improving those educational results.

Some Dangers of Scientific Measurement to be Avoided

(Abstract)

FRANK W. BALLOU, *Superintendent of Schools, Washington, D.C.*

WHAT I shall have to say on the dangers of scientific measurement has grown out of my experience and is intimately related to my address of Friday afternoon ["Improving Instruction Thru Educational Measurement"]. In fact, these two addresses supplement one another. The one on Friday afternoon showed what was accomplished over a period of years, thru educational measurement; this talk will concern itself with some of the dangers of which all school people must take cognizance if these dangers are to be avoided.

At the outset I want to apologize for the dogmatic method which I shall follow in this talk. I follow that method because it makes for directness, because I can thereby cover more ground, and also because I want to close on time, so that time shall be left for discussion following my talk. I shall discuss four dangers, the first of which is:

1. DANGER THAT RESULTS OF MEASUREMENT WILL NOT BE UNDERSTOOD BY THE TEACHERS

When I began work in Boston, the Courtis Standard Tests had been given several times. A very intelligent teacher stated that her pupils had taken the test three times, but she could see no change in them. This teacher seemed to feel that the tests were like medicine, and that the children should improve in their educational work after each dose. The teacher must be taught the purpose and use of the results of educational tests if they are to make appropriate use of them. I need not say to this group of people that there is nothing magical about the tests. The value of the tests depends upon the intelligent use by teachers and school officers of the information which the tests furnish them.

Educational measurement has necessitated our familiarizing ourselves with new terms and indeed a new language. The terms "median", "mode", "intelligence quotient" (i. q.) and many others must now become a part of our professional language. Teachers must be familiar with these terms if they are to read intelligently professional journals.

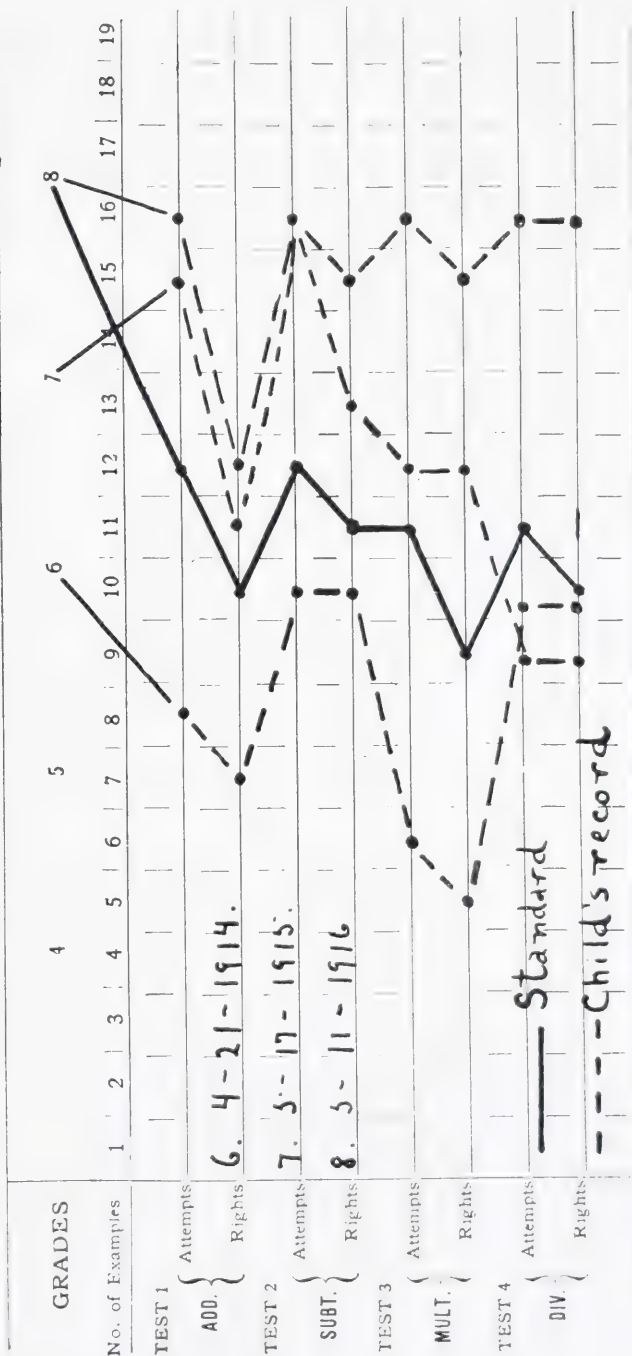
Not only must teachers be familiar with the new language of educational measurement but they must also be able to interpret correctly charts, graphs, and other means used to represent the results secured thru educational measurement. I present herewith some charts about which I shall ask some questions and make some comments to illustrate the necessity of teachers being familiar with such charts.

This record shows that this girl appears to have made unusual progress in drill work during the eighth year. The question is whether the child was not overdrilled. Only the teacher who taught the child can answer this question. For an eighth-grade pupil to improve in drill work from the ability represented by 6 examples attempted and none of them right to 19 attempted and 19 right, represents great improvement. Was this improvement secured at the sacrifice of other studies?

This boy, thirteen years old, who graduated from the eighth grade was deficient in addition and division particularly. The teachers must be trained to use this information and should be able to explain why the pupil can be trained to do satisfactory work in subtraction, somewhat less satisfactory work in multiplication, but be unable to do division.

RECORD OF ARITHMETIC TESTS.

Name a w girl Age at first trial 12
(On last birthday)
(Boy or Girl)



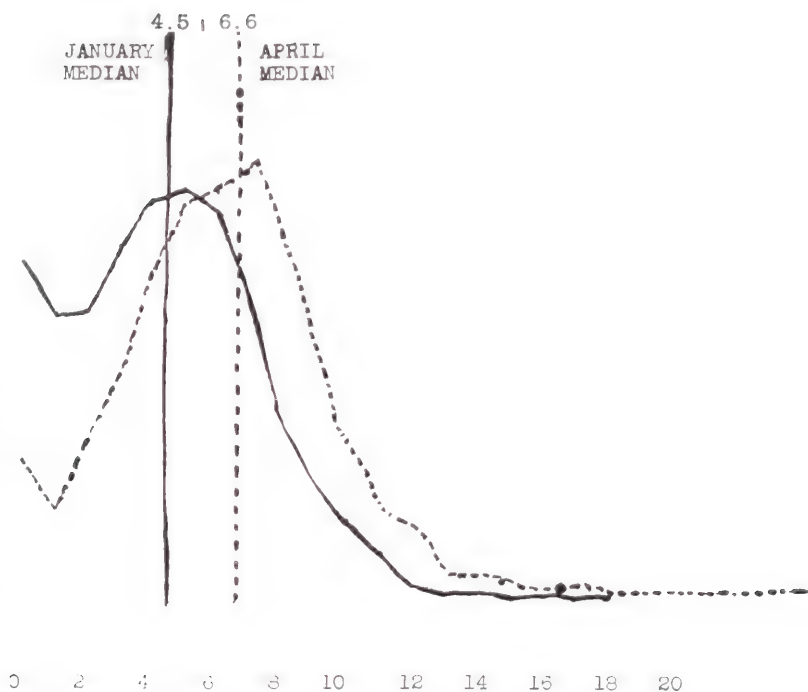
Legend.— On each horizontal line after the word Attempts, place a dot under the figure representing the number of examples you have tried, and on the line after the word Rights, place a dot under the figure representing the number of examples that you had right. Beginning with the figure at the top that represents the grade draw a line through all the dots that you have made.

The record of A. W. is stated in graphic form. Teachers must be trained to interpret such graphs, since they are becoming increasingly important in our educational literature.

- 6 -

DISTRIBUTION OF JANUARY AND APRIL SCORES COMPARED

SUBTRACTION RIGHTS GRADE V.



The above graph shows that the children of the fifth grade advanced from a median of 4.5 to a median of 6.6, from January until April. The teachers must understand the significance of this. Not every child made progress. Some children undoubtedly did less well in April than in January. The median represents the combined progress of all individuals.

Tabulation Sheet of a Class which Approximates the Established Standards in Speed, Accuracy and Efficiency.
Grade VIII, Division, May, 1916.

SCORE IN NUMBER OF EXAMPLES ATTEMPTED.																										
SCORE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	TOTAL
100									1	4	3		2	1	1											12
90												1	1	5		2	2									11
80					1	2			3	2	1	2						1								12
70								1			1															2
60						1			1	1	1		1				1									5
50																										0
0 to 49					1							1														2
Total						2	3	1	5	7	7	3	8	1	1	3		3								44

Median scores of this class: Speed, 11.6. Accuracy, 90.9 per cent. Efficiency, 40.9 per cent.
Standards: Speed, 11. Accuracy, 90 per cent. Efficiency, 39.7 per cent.

This is the record of the class which reaches approximately the standards established for the school system. It is to be noted that even tho this class approximates the standard, nevertheless there are many children who need special attention. The median is a class measure. Individuals in that class must be considered. For example, one pupil attempts only 5 problems and gets less than half of them correct. Three pupils attempted as few as 6 examples, with an accuracy of from 60 to 80 per cent. Such cases must receive consideration, even tho the class record is satisfactory.

2. DANGER THAT TESTS WILL BE SO ORGANIZED THAT THE ABILITIES OF PUPILS WILL NOT BE ACCURATELY REPRESENTED BY THE STATEMENT OF RESULTS

Obviously this is a danger which must be of chief concern to those who compare achievement tests or intelligence tests. Nevertheless it is of such importance that it is worth while to give some attention to it on this occasion.

To illustrate, I select the Courtis Standard Tests in arithmetic. You will recall that each problem in the Courtis Standard Test in addition consists of 3 columns of figures, with 9 figures in a column. Let me recall for you also that the standard of achievement based upon these tests for the city of Boston was 12 problems attempted with 10 problems correctly solved, or an accuracy of 83.3 per cent.

The various possibilities with respect to accuracy are shown in the following tabulation:

TABLE OF ILLUSTRATIONS

	Attempts	Rights	Per Cent. of Accuracy
STANDARD	12	10	83.3%
	10	8	80.0%
	11	9	81.8%
	12	10	83.3%
	13	11	84.6%
	14	12	85.7%
	15	13	86.6%

Let us examine the actual facts with respect to the accuracy with which pupils do their work in reaching the standard established. To solve each one of the problems in addition a pupil must make 26 computations; 8 computations in the first column and 9 computations in each of the two following columns, owing to the number carried. If he solves the 12 represented by the standard, he must therefore make 312 computations (12x26). If the candidate made one error in each of 2

problems his accuracy is actually 310 computations out of 312, or a very high percentage of accuracy. The computed percentage of accuracy of 83.3 per cent, based on 12 problems attempted and 10 right, represents the pupil's accuracy only in case he failed in every one of the 52 computations in the 2 examples which were marked incorrect.

3. DANGER THAT TESTING WILL MISDIRECT TEACHING

If testing is for the purpose of revealing facts to the teachers to the end that the teachers' instruction may be improved, then educational measurement is of primary concern to the teacher. The teacher must be expected to be influenced by the results of the testing. Teachers, as a rule, undertake to meet whatever standards are established for their respective classes. Herein lies the great advantage as well as the disadvantage of testing. It is an advantage in case the standards are correct. It may become a very great disadvantage if the results of the testing or the standards misdirect the teachers. A few observations will make this point clear.

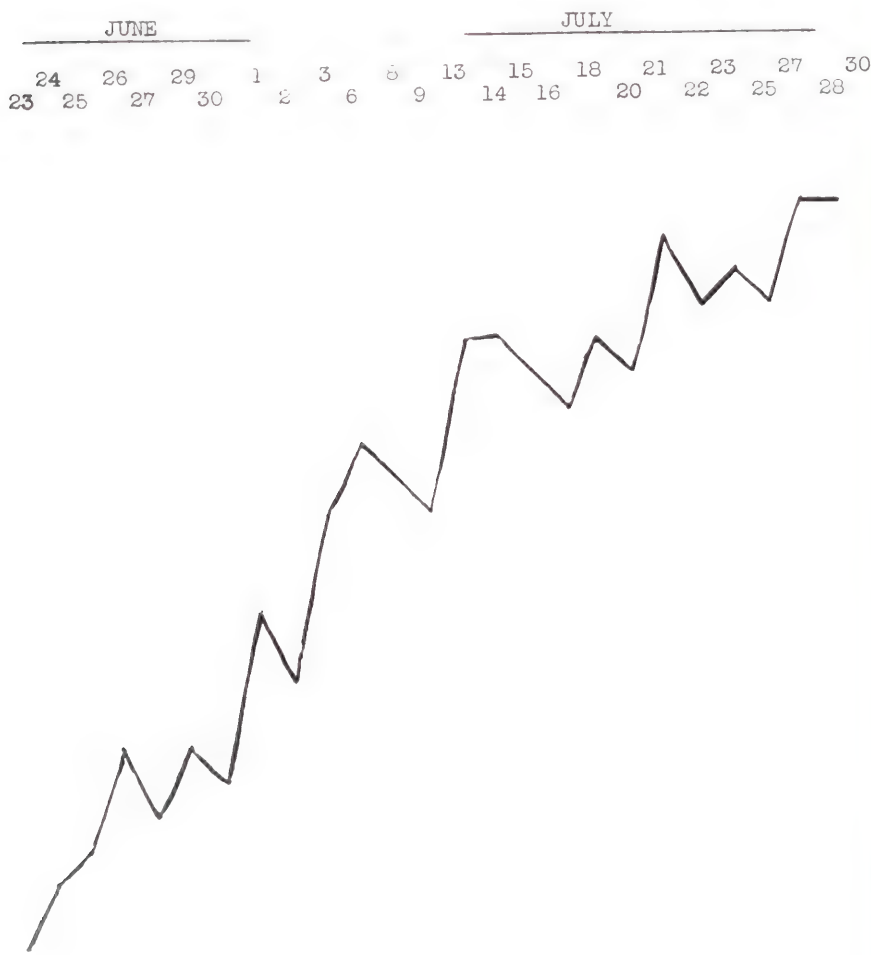
For the most part, educational measurement thus far has had to do with the formal aspects of subjects tested. Various phases of arithmetic have been accurately measured. Standard tests in geography have had to do largely with place geography. The formal side of English composition has been measured, but less success has been obtained in the content side of these and other subjects.

When the Department of Educational Measurement was organized in Boston it attempted to establish lists of words for spelling which were most difficult and which every pupil should know how to spell. At the beginning the number of words assigned for each grade was comparatively small. At the end of each year pupils were tested with these words. In their zeal the teachers had taught these words so thoroly that pupils spelled them with a high degree of accuracy. As the number of words for each grade was increased teachers were unable to give such intensive training in each word. The result is that the standard of accuracy established for the words originally carried in the lists for the various grades has not been reached in later years. Undoubtedly too much time and attention was devoted to spelling in those early days.

When the Harvard-Newton scales for the measurement of English composition were being prepared it was the intention to provide a scale for each of the five forms of English composition, namely: description, narration, exposition, argumentation, and reproduction. After careful consideration those preparing the composition scales concluded to omit reproduction since it is a form of composition that should probably not be further emphasized in our schools. It was felt that reproduction is too frequently resorted to by the teachers, that it is not the best form of training. Inevitably teachers would feel that it was equally important with other forms of composition should a measuring scale be provided for it, hence this composition scale was not printed for distribution.

There is another aspect of the measuring of English composition which I desire to allude to. The most usual form of composition scale

is one which contains a series of compositions by which the compositions of children of elementary school and high school shall be measured. I desire to record my opinion that poor compositions in grade eight, that for example may be rated 60 per cent, are not fair measures of the work done in the grades preceding grade eight. The composition set up as the best composition in a scale should be an ideal towards which the pupil may work as well as the measure of the quality of their efforts. A 60 per cent composition written by an eighth grade pupil cannot possibly be held up as an ideal toward which children in the lower grades may work.



4. DANGER THAT THE RESULTS OF TESTING WILL BE GIVEN TOO GREAT WEIGHT

In view of my address on Friday afternoon, I feel that there is no danger of my being misunderstood when I say that the results of testing work may be given too great weight. I consider those results invaluable; however, while the results of testing constitute important data they are not exclusive data. Scholastic marks are still of importance. Pupils' interests and attitudes must likewise be taken into consideration along with the results of testing, if school officials and teachers are to exercise wise direction of a pupil's education.

Intelligence tests and accomplishment tests must supplement one another. The results of these two tests, together with the knowledge of the pupil and his ability which teachers and officers have heretofore possessed, will form a comprehensive basis on which to plan a pupil's education.

There is danger that the results of the testing may be given too great weight if only one test of a pupil or a class is used as the basis of judgment. As an illustration I insert here a graph which shows the progress of a class exercised with drills and tested daily over a period of thirty school days.

The graph indicates that the progress of the class is not uniformly upward, but that, on the contrary, on some days the class as a whole does better than it does on other days. Its progress in general is upward. Care must be taken in the use of such results, because when the class is tested it may be that the class is (*a*) on the mountain peak, (*b*) in the valley, or (*c*) on a plateau. The results ought to be considered accordingly.

CONCLUSIONS

I have indicated four possible dangers in educational measurement. In conclusion I desire to indicate how those dangers can be avoided.

1. The danger of a misunderstanding of testing on the part of the teachers can be avoided by the instruction and training which is now a part of the preparation of future teachers. Teachers now in the service can be provided instruction by their officers and can receive invaluable instruction thru conferences like this.

2. The danger that the results from a test will not represent the abilities of the pupils can be avoided by the exercising of more care in the preparation of tests. This is a danger which the comparatively few people preparing tests must undertake to avoid.

3. The danger that the tests will misdirect teaching can be avoided thru the extension of testing work to cover the whole field of instruction; by the exercising of care in organizing and carrying on testing work so that the attention of teachers will be spread over a large field of instruction; and finally thru teachers trained in the value and significance of educational measurement.

4. The danger that the tests will be given too great weight will be avoided by a better understanding on the part of teachers and officers of the place of such testing in a general program of educational betterment.

Education and Our Responsibility for Its Improvement

FRANK W. BALLOU, *Superintendent of Schools, Washington, D.C.*

EDUCATION in the United States today faces one of the most serious crises in its history. The war revealed the inadequacy of some parts of our educational program. The war likewise created conditions in our school system which it will take at least one generation, and perhaps many generations, to eradicate. An enlarged and a more intensive educational program is a national necessity. And yet the financial resources provided for public education are not sufficient to do more than to maintain the educational standards existing before the war. The policy of financial retrenchment in state and nation threatens to cheapen the education of boys and girls now in school.

REVELATIONS AND RESULTS OF THE WAR

The financial burdens resulting from the war are large. However great our financial burdens may now be, they should not be allowed to blind us to the necessity of placing adequate financial resources at the disposal of the state for the appropriate education of the youth of America now found in our elementary schools, secondary schools, colleges, and universities. As far as possible, the burden of war debts should be borne by the present generation. Financial obligations incurred by the present generation should not be passed on to future generations. There is a grave danger that the financial burdens of the war are likely to be visited on the pupils now in our public schools thru reduced educational opportunities provided for them.

For the most part, the young men inducted into the American Expeditionary Forces were the product of the public school system. No one can successfully assail their courage, pluck, alertness of mind, and other personal qualities which go to make efficient citizens in times of peace or of war. They met every test of good soldiers, as they had previously given promise of meeting every requirement of good citizens. The nation owes a debt of profound gratitude which is bound to increase as the years go by, not only to those who sleep in the poppy fields of France, but also to those who have returned home to assume the responsibilities of citizenship in popular government.

The less favorable conditions among our youthful citizens revealed by the war must likewise be faced and cannot be lightly passed by. While we extol the qualities of those who served the nation, we must also give attention to those young men of draft age who were rejected or put into limited service. The army educational tests showed that 29.4 per cent of the men included in the first draft were unable to pass successfully a simple test of intelligent reading and legible writing.

The draft of June, 1917, showed nearly a hundred thousand illiterates or those who were approximately illiterates. The majority of these illiterates were native-born. They were unable to serve their country in time of war because of their illiteracy.

Moreover, the records show that of the 3,208,000 men included in the first draft and physically examined, 949,000 or 29.6 per cent were totally or partially disqualified for military duty for physical reasons. In addition, the last Federal census shows that there were 4,931,905 illiterates in the United States. Of these, 3,084,748 were native-born.

INADEQUACY NOT INCOMPETENCY OF EDUCATION

These figures on illiteracy among our native-born young men are not so much evidence of failure to do well what the school system undertook to do but rather failure to make the educational system as comprehensive as it ought to be. These facts reveal the incompleteness of public education rather than its incompetency.

When one reviews the conditions in the educational system of America, the revelations of the war do not appear startling. On the contrary, they are the logical results of conditions existing in our system of education. The following figures represent conditions which contribute to the ineffectiveness of American education:

Of the 600,000 teachers in American elementary and secondary schools, 30,000 have only an eighth-grade education or less. One-fourth of the 600,000 teachers have had only two years of high school work or less. Only one-fifth of the 600,000 teachers possess the standard of professional preparation recognized as a minimum by nations generally. One-fourth of these 600,000 teachers have had less than two years of experience and one-half of them have had less than four years of experience.

The typical rural school teacher in America has had less than two years of high school work, is under twenty-one years of age, has had less than two years of experience in teaching and is at least twenty miles from a superintendent or supervisor who visits her only once a year.

The inadequacy of American education is to be accounted for, in a large measure, by the conditions indicated. These conditions contributed to or permitted the educational defects among the men drafted for the Great War. These conditions and others similar to them made possible the near illiteracy and physical defects revealed in the drafts. No improvement can be made in overcoming illiteracy or in improving the physical condition of our boys and girls until the conditions which made possible the revelations of the draft shall have been improved.

EFFECTS OF THE WAR ON SCHOOL SYSTEM

The school as a social institution would naturally not escape from participation in the war activities or certain effects brought about as a result of the great conflict. Many teachers and school officers were called into service directly connected with the war. Teachers and officers who remained at their school posts devoted themselves not only to

their school work but also to many and varied war activities outside of the classroom. Moreover, pupils were called upon to participate in the great variety of war activities, during the time ordinarily devoted directly to more direct educational work. The school system as a whole contributed in many ways directly and indirectly to the prosecution of the war.

While the school system is apparently gradually returning to normal conditions and is devoting itself to its pre-war educational progress, there is one outstanding result of the war which is likely to be overlooked. Most school systems replaced the teachers and officers entering the war service with persons far below the professional and educational qualifications of those who left. Thus far the supply of competently trained teachers has not been sufficiently large to replace those persons in the school system whose professional qualifications and teaching skill are not up to to the American standard. Even if salary conditions which prevailed preceding the war are re-established it will take at least a generation to re-establish the professional standing of the teaching profession in America. If the salaries are not raised to a level which will represent a purchasing power equal to the purchasing power of teachers' salaries before the war, the professional standards of the teaching profession must continue to be lower for more than a generation.

The war brought all schoolhouse construction to a standstill. Neither labor or building material was available for the carrying on of the usual amount of building to meet the needs of our growing systems of public education. As a result, the school buildings of the nation are wholly inadequate to house satisfactorily the pupils who are annually seeking education in elementary schools, high schools, colleges, and universities.

THE PRESENT SITUATION

During the war it took the undivided attention of educators generally to keep the schools running at all. The shortage of teachers was universal. The lack of school buildings in which to house the pupils gradually became acute. Other administrative difficulties have so completely occupied the time of the school men and women that little or no thought has been devoted to educational aims, methods of teaching, and educational results. The discussion of educational principles is almost completely left out of many of our educational meetings, and discussion of the schoolhouses, the school budget, and the shortage of teachers has taken its place.

While better buildings make better education a possibility, it takes more than a good school building to make an educational institution. The longer adequate appropriations for buildings are delayed, the longer will the discussion of educational principles be postponed, because the latter must await the former. And better educational results must likewise await both.

Our elementary schools thruout the cities of the country are greatly overcrowded. Part-time prevails in astonishing proportions. One-room portable schools are being used instead of well-equipped permanent

buildings. Classes have been increased in size beyond reason. The educational program has been curtailed in many ways. The so-called Gary plan has been tried as a means of accommodating a larger number of pupils than the buildings were constructed to accommodate. Other plans of various kinds which make our schools into factories or mills instead of efficient educational institutions have been put into operation in an attempt to educate the masses of youth of the present generation.

Larger and larger numbers of our boys and girls are going to high school. This should be our glory, but it is in reality an embarrassment. High schools have been forced to accommodate pupils beyond their capacities. Various kinds of double-shift programs have been devised to make at least the attendance of pupils a possibility. Some day when these same boys and girls shall have become citizens I fear they will be discussing and condemning the conditions which we permitted to stand between them and an adequate education.

Colleges and universities are likewise crowded. There is a marked difference between the policy followed in the public schools and the policy followed in the colleges and universities. The public school officials must make room for all pupils who apply. On the contrary, when the enrollment of a college or university reaches the capacity of the institution, frequently the doors are closed and admission denied.

When I was associated with the University of Cincinnati more than ten years ago, most colleges and universities were seeking students. They were competing with one another to increase their enrollments. Today the situation is the reverse. For some time, many colleges for women have limited their enrollment.

Recently there has developed much public discussion which is the direct outgrowth of the large enrollment in our colleges and universities. Instead of undertaking to provide suitable collegiate and university training for the men and women who are seeking it, the present agitation has to do with reasons why men and women should not go to college.

The present discussion centers around the statement of President Ernest M. Hopkins of Dartmouth, on September 21, when he made the general assertion that "Too many men are going to college." On that occasion President Hopkins is reported to have said:

"The opportunities for securing an education by way of the college course are definitely a privilege and not at all a universal right. The funds available for appropriation to the uses of institutions of higher learning are not limitless and cannot be made so whether their origin be sought in the resources of public taxation or in the securable benefactions for the enhancing of private endowments. It consequently becomes essential that a working theory be sought that will operate with some degree of accuracy to define the individuals who shall make up the group to whom, in justice to the public good, the privilege shall be extended, and to specify those from whom the privilege should be withheld."

Following the publication of those views of President Hopkins, Chancellor E. E. Brown of New York University has said:

"There are too many college students only if their mere numbers prevent us from giving a fit training to the destined leaders of men within the next generation, but we cannot admit this without confessing a certain bankruptcy of inventive resourcefulness, a limitation which, as thinking men, we cannot accept."

President Sidney Mezes of the College of the City of New York says:

"If President Hopkins means to say that the country is getting too many college-trained men, I must disagree with him. The country cannot have too many men who are well-trained. But I think he was talking of those persons who are actually in college who ought not to be there."

I do not desire to enter into a discussion of the merits of the proposition advanced by President Hopkins. I cite it only as an illustration of what I believe to be an unworthy attempt to justify limiting the number of young men and women going to college. I believe that the time will come when every college and university will feel as complete a responsibility for enrolling all young men and women who desire to pursue collegiate education as does the public school system at the present time. Collegiate education should be as universally open to young men and women as is high school education. Moreover, I believe that the time will come when colleges and universities will conduct a careful self-examination to ascertain whether the program of education provided in our colleges is as diversified as are the social needs for which colleges should train men and women, and as varied as are the capacities, interests, and future careers of the young men and women who enter those institutions. I believe that such a self-examination will reveal that many young men and women are considered unsuccessful in their college work only because they do not fit into a narrowly conceived collegiate program, rather than because they do not possess the ability to pursue proper collegiate training which will fit them for their future careers as useful citizens.

ENLARGED EDUCATIONAL PROGRAM A NECESSITY

The revelations of the war and the effects of the war have made increased demands on the schools for education of a better quality and for an educational program broader in scope. Increased attention has been given to the problem of Americanization of foreigners. Renewed interest has been exhibited in the passage and enforcement of compulsory education laws to the end that native-born boys and girls shall not grow up illiterate. There is a tendency toward increasing the length of the school year. Summer schools and evening schools are being extended to provide additional educational facilities.

Increased attention is likewise being given to physical education. Medical examination of pupils is a part of the educational program to the end that the physical defects of the boys and girls may be revealed and corrected during the progress of pupils thru the schools. Legislation has been passed in several states prescribing physical education as a part of the school program.

EXPENDITURES VS. POLICY OF ECONOMY

These enlarged educational opportunities and these extensions of the educational program involve larger expenditures of money. This demand for increased expenditures for public education appears to fly in the face of a public demand for lowering the cost of government generally. To contribute to the solution of the problem a comprehensive national investigation of financing public education is under way. It is the purpose of this inquiry to reveal, if possible, additional sources of revenue for public education. Efforts have likewise been made to secure financial support from the national government to the end that states may be assisted in establishing a larger and more comprehensive program of public education.

In the last analysis any educational program must be reduced to a money basis. Money alone will not make a school system efficient, but a school system cannot be made efficient without adequate appropriations. Whether a system of public education of the future shall produce better results than the school system of the past depends more largely upon appropriations for public education than upon any other single factor.

In an address before the Department of Superintendence of the National Education Association, in Chicago, in 1922, Mr. Will C. Wood, state superintendent of public instruction of California, undertook to answer the question, "What has this country reason to expect by way of culture, efficiency, and good citizenship in return for the cost of operating public education?" I quote two paragraphs from Superintendent Wood's address:

"The natural point of departure in this discussion is the amount we are now expending for the elementary and secondary schools of our country. The figures for the year ended June 30, 1920, show a total expenditure for the public schools by all states of a trifle less than \$900,000,000. With the growth of population and the upward trend of wages that persisted in the early part of the last fiscal year, we may safely assume that the total expenditure for public education during the year ended June 30, 1921, was between \$1,000,000,000 and \$1,100,000,000, which is approximately \$10 for each man, woman, and child in the country. Let us compare this expenditure for that year with the total expenditure of 1911, a decade previous. The total expenditure during that year was \$446,727,000, or \$4.76 for each unit of population. The federal Department of Labor is authority for the statement that during the year 1921 the purchasing power of a dollar was less than half its purchasing power in 1911. In other words, the \$10 per capita expended for public education in 1921 could purchase no more than the \$4.76 per capita expended in 1911. The so-called increases in school expenditures during the last decade have therefore not been increases at all. School expenditures measured in purchasing-power are no greater today than they were before the war. After all the agitation and enthusiasm for education growing out of awakened recognition of the nation's dependence upon and duty toward the youth it had drafted for its defense, we find that expenditure for schools has barely kept up to the pre-war level.

The country as a whole is spending for education in 1922 in less proportion of its total income than in 1911. It must therefore be observed that the schools are guiltless of the greed of the profiteer.

"What in the way of culture, efficiency, and good citizenship has this country reason to expect in 1922 in return for the cost of operating the public schools? The answer is simple and direct: inasmuch as it is spending no more of purchasing power upon the schools in 1922 than it was expending in 1911, it has reason to expect no more by way of culture, efficiency, and good citizenship than it secured by the expenditures of 1911. It has reason to expect no more, but as a matter of fact it is getting more, the increase being chargeable to the devotion of the great body of teachers and school administrators to the cause of better America."

I am not an alarmist. I do not approve of exaggeration as a means of emphasis. I believe in the fundamental soundness of our program of American education. Nevertheless, the educational program of pre-war period, and still less the educational program made obviously necessary by the revelations of the war, cannot be financed with the same or only slightly increased appropriations, when the dollar is buying less than one-half as much education as it purchased before the war.

OUR RESPONSIBILITY

School men and women in elementary schools, high schools, colleges, and universities must assume a larger leadership in their respective communities, in their states, and in the nation. The teaching profession has never begun to realize the possibilities which are bound up in its members, in its prestige in state and nation, and in its capacity for public service. Thru the National Education Association with its 100,000 members, thru the various state associations affiliated therewith, and thru the still larger number of city and community associations, the teaching profession has an unprecedented opportunity for the exercise of constructive educational leadership in community, state, and nation.

We must always remember that at any time we are only one generation removed from a generation of ignorant people. Close educational institutions for a brief period of one generation and the next generation must be illiterate. Each generation must provide the educational facilities necessary to prepare the next generation of men and women to sustain and improve our democratic institutions. Whenever a democratic nation is financially unable to provide an adequate education for its future citizens, it is not only financially bankrupt but it will soon become bankrupt educationally, politically, and morally.

Let us, as members of the great profession of teaching, consecrate ourselves to the task of creating an intelligent public opinion concerning the importance of education in a democracy, to the end that the generation of young people now in our schools, colleges, and universities shall not be deprived of an adequate education, which has gradually been established as the birthright of every boy and girl in America.

Selling Educational Research to a Large City School System

ROBERT H. LANE, *Assistant Superintendent of Schools, Los Angeles, Calif.*

I THINK I should say by way of preface that I have come before you under false pretenses. I am not a research man either by training or experience—I am simply a salesman, comparable to those knights of the road who sell you anything from bonds to battleships, or possibly more nearly akin to their humbler brothers who peddle soap from door to door or who lie in wait for you at the office with the world's best literature complete in thirty volumes, a monumental work which no intelligent man could afford to be without.

I warn you also, that this is not a formal address upon educational research. It is merely the plain, unvarnished tale of how two men attempted single-handed (an expression worthy of Sir Boyle Roche, by the way) to sell educational research to a large city school system which had gotten along without it, thank you, for a goodly number of years. Such a form of exposition has at once its virtues and its faults—you will have to pardon a certain diffuseness, a wearisome repetition, and a multiplicity of personal pronouns. On the other hand, you will gain by hearing a recital of actual occurrences, vivid, concrete, and I hope interesting and helpful to all of you who, in one form or another, are selling research to individuals or to groups of persons interested in education.

On the first day of May, 1917, Dr. Arthur Sutherland and myself were transferred to the superintendent's office, one of us as director of the division of psychology and the other as director of educational research. Dr. Sutherland's instructions were brief and definite—he was to locate the feeble-minded children in the city schools, segregate them into groups, and provide for them instruction by competent teachers. My instructions were equally brief but hardly as definite. The director of educational research was to ascertain periodically if the elementary pupils in our city schools were able to compete on equal terms with pupils of other cities in the matter of educational group tests standardized at that time. If our pupils were able to present a creditable showing, well and good; if not, the director was to devise methods and devices whereby the deficiencies could be most quickly overcome. You will notice at once the significant fact that the whole purpose of research was held to be administrative. It appeared to be simply good policy and good advertising to have the city schools as a whole make as good a record in the educational world as the schools of other systems. That educational research could be intimately connected up with the needs of individual pupils, that it could be a mighty lever to move the dead-weight of maladjusted and retarded pupils, was apparently not even dreamed of.

It seemed to both of us as we planned our work that we could follow one of two possible courses: (a) We could give thru the machinery of the superintendent's office a series of city-wide group tests, using the kind offices of volunteer workers from the local universities and high schools in scoring and tabulating the results of the tests. Such a plan would bring immediate returns in the shape of records covering all of our schools with the expenditure of comparatively little time. From the administrative standpoint this was the ideal course of action. It had the fatal defect of touching the immediate needs and interests and daily tasks of the classroom teacher not at all, except in a most superficial way. Or, (b), we could work individually with principals and teachers, giving our tests in person, and scoring them and interpreting them so as to make constructive diagnosis possible. We realized from the first that such a plan would be painfully slow and that we would have little to show in the way of results for a long time to come, but since we were building for the future, since we had all the years ahead of us to work in, and above all, since we had faith in the willingness and ability of our people to grasp this new, strange device in education, once it was fairly presented to them, we decided upon a salesmanship campaign whereby educational research should be "sold" in person to the principals and teachers of our schools. May I point out to you, therefore, that by and large our sales campaign passed thru four distinct phases, the first of which I shall call

I. The House-to-House Canvassing Phase. *Heteme aqui ya fuera de Oviedo, camino de Penafior,*" says Gil Blas at the beginning of his travels. "Behold me therefore just outside Oviedo, on the way to Penafior." Behold the speaker, therefore, one bright day in May armed with a brief case packed with the popular tests of that day descend upon a large elementary school whose principal desired to know what this new educational device might be and how it worked. By chance I found a school composed largely of Jewish children who were intensely interested in arithmetic and who responded to the Courtis Tests in Arithmetic with great enthusiasm. The tests were given in person and teachers and principal invited to watch the proceedings. The papers were scored and the results tabulated on the spot, and some attempt was made to interpret the variations in class and individual ability which were plain to behold. The next day, another school was visited where reading was popular and the then Kansas Silent Reading Tests proved the entering wedge. For several weeks the process was repeated until a growing interest in our tests led to several invitations to visit schools whose principals and teachers had not been conspicuously friendly to the test idea at the inception of the campaign but who had later become interested from reports emanating from the schools originally visited. It should be noted that the salesman evolved thru experiment a daily schedule which proved fairly satisfactory. Each morning a school was visited, tests given as far thru the classes as time would permit, and the afternoon of the same day was devoted to marking papers, tabulating results, and to preparing a crude report which was later submitted in person to the principal and his teachers so as to es-

establish a basis for profitable discussion. By the end of the term, about one-half of our schools had been visited, and while the work in any one school was painfully superficial and inadequate, the idea was beginning to "sell" and there was a growing demand for our "goods." Occasionally mimeographed reports of combined results were prepared showing that in the main our elementary pupils were conspicuously behind in some subjects compared with children elsewhere while presenting a favorable result in other subjects.

The steady increase in our school population at this time and the consequent necessity for hastening our progress thru the schools led to the second phase of our campaign which I shall call

II. The Instructional Classes Phase. By way of experiment, we inserted in our local school journal a modest notice to the effect that we would, at the opening of the next school term, inaugurate a class in educational measurement for elementary school principals who desired to know enough of this new idea in education to experiment in their schools. We fixed the membership of the class at twenty students, not because we wished really to restrict the size of the class but because we had no idea how many replies would be received in answer to our invitation, if any at all. In other words we acted on the well-known principle laid down by Charles Dickens that one of the best ways to get people to attend a performance is to impress upon them that it is impossible to get in. To our surprise we enrolled twenty-four principals in this first venture of ours and carried them thru a five months' course, meeting once a week after school. The method of instruction was simplicity itself—our principals were told to consider themselves elementary pupils and tests were given to them just as we gave them to our boys and girls. Then the papers were marked and the results tabulated on the blackboard and individual difficulties explained away. Diagnostic methods were taken up, and the members of the class were shown in a crude way how to use the results of tests to locate deficiencies in instruction and administration. As the available textbooks proved too difficult, a simplified drill in educational statistics was devised to insure ease and accuracy in handling data. At the end of the first semester, a call was sent out for a new volunteer class for the succeeding semester, and as at first, the membership was set at twenty pupils. We were gratified to enroll not twenty but seventy-seven principals in this new class which repeated the previous course, while the earlier class continued as an advanced section taking up group intelligence tests which were just then coming into prominence. I should like to point out that we were apparently "selling the idea" not only to our own people but to the local universities as well. At the close of that year the demand for advanced instruction in educational measurement became so pronounced that both universities offered a variety of courses both in summer session and after-school classes, thus relieving us of any further responsibility in the matter. The universities greatly strengthened our hands not only by providing most competent instructors but by offering university credit toward a degree upon these courses.

I now come to a phase of our campaign which is difficult to describe

as we passed thru a transition stage common to city school systems at periodic intervals. For want of a better name I shall call it

III. Salesmanship Thru Special Rooms and Special Teachers. In January, 1920, I was transferred to the superintendent's office and Dr. Sutherland was given the new title "director of the division of psychology and educational research", facing the monumental task of doing single-handed the work formerly divided between two men. Work in the former division of psychology had proceeded briskly since 1917 and three types of special rooms were devised to remedy existing inequalities: (a) development rooms for definitely feeble-minded children, (b) opportunity rooms for both border-line cases and an occasional super-bright child, (c) adjustment rooms. The adjustment room as developed at Los Angeles calls for more than passing notice. There are four outstanding characteristics of the adjustment room. (1) Entrants must be normal children who have fallen behind in their school work thru illness, absence, defective teaching, or other contributing cause, but they may not be subnormal children or disciplinary cases or physically defective. In other words, only such children may be entered in an adjustment room as seem to be likely material for educational diagnosis and capable of profiting by the peculiar type of instruction devised for these rooms. (2) Pupils remain in the adjustment room for a limited period—usually about sixty working days. (3) Adjustment material consisting of mimeographed exercises, projects, tests, and practice sheets from an informal textbook while the official textbooks are used in a supplementary capacity. This adjustment material is based on twelve levels of attainment from low first to high sixth grades, each level representing roughly a half-year's progress on the pupil's part. (4) The unique feature of the adjustment room lies in the fact that the pupil pulls the load while the teacher acts in an advisory capacity. Provision for self-initiative is made thru the material itself and thru fixing on the pupil responsibility for his own progress.

During the transition period to which I have referred, pupils were admitted to adjustment rooms and development rooms only after tests were given either thruout the several schools or in part and after eligible lists of pupils had been prepared. I wish you to notice the administrative machinery whereby this testing was done. The city was divided into several districts, and over each of these was placed a supervisor whose specific job it was to (a) give such tests thruout the several schools as would determine what types of special rooms were most needed, (b) prepare eligible lists of pupils for both adjustment and development rooms, (c) help new adjustment teachers to organize their work, (d) supervise such rooms after organization until they become autonomous; i.e. until the teacher in charge could perform her work so efficiently that she needed only casual and occasional supervision. Please notice that in effect this scheme placed the research program in so far as tests were concerned in charge of a small group of supervising teachers and removed it almost completely from the jurisdiction of the principals and regular classrooms. I need not point out that such a form of salesmanship had one outstanding virtue and one almost fatal defect as you will

see. It had the outstanding virtue of centralizing the research program in a small, easily-controlled body of specialists who reported directly to their administrative head so that a standardized procedure was easily worked out. The fatal defect of the whole scheme lay in the fact that principal and regular teachers were practically ignored. The teacher-supervisor came to a school, gave tests, prepared eligible lists, organized special rooms, and reported her findings directly to her chief. The special room teacher after her work was organized also reported directly to the head of the department, and the principal's share in the whole educational drama was merely that of a bystander. This is no reflection on the efficient and often remarkable work done by the department in taking care of the misfit, but it led to a growing unrest among the principals who woke up one morning to find the testing program completely removed from their hands. Such was our situation up to the beginning of the last school year. This brings me to the fourth and last phase of our campaign which I shall call

IV. The School Counselor Movement. I am indebted for this solution to our difficulties indirectly to my good friend, Dr. Virgil Dickson of Oakland, and directly to the casual visit of an Oakland teacher who, in discussing educational research, said that in Oakland a plan had been devised whereby a teacher in each of several large elementary schools had been partially relieved of her classroom duties so as to permit her to give tests and assist in grading and classifying the schools. This plan seemed to me to secure both those desirable ends which hitherto had proved to be unattainable simultaneously: (a) to place the testing program under such close supervision as to make it most effective, and (b) to give principals and teachers such a large share in the work as to enlist their hearty sympathy and support. Dr. Sutherland proved most helpful in planning to make the experiment in our own system, and as a beginning three elementary schools were selected for the purpose, the choice being finally made in schools where at least one teacher in each school had taken a rather unusual interest in experimental education. Please notice that we did not send a teacher from the outside to each school; we selected a teacher in each case who had been in her school for some time, who knew local conditions and needs, and who had the confidence and respect of her fellow-teachers and her principal. I should like to point out also that each of these three schools had an exceptional principal, one who was interested in our campaign and who could intelligently direct the work. These three "school counselors" as we shall hereafter call them were given the following instructions:

"Your job is to assist the principal to regrade and reclassify his school to the end that children shall be grouped together who are at or near the same educational level and who are presumably capable of making about equal progress in their work. When your survey is complete, go over it with your principal, and make such regroupings as will prove mutually satisfactory to both teachers and principal. If a teacher's judgment of a pupil disagrees radically with the results of your tests, give the pupil the benefit of the doubt and stand by the teacher's judgment until she is clearly proved to be wrong. Report your results to

Dr. Sutherland's office and obtain his advice on knotty problems which may arise from time to time."

One of the three school counselors had just returned from a year's work at Teacher's College and desired to follow the plan of reclassifying devised by Dr. McCall. We asked the two other counselors to follow the same procedure so as to unify our work, and as a result the survey tests consisted of the Woody McCall Reading, Thorndike McCall Arithmetic, National Intelligence Test, and a selected group of words from the Ayres' Spelling Scale. From the results of these tests, together with the teacher's judgment, pupils were reassigned to the several grades and divided into X, Y, and Z groups for instructional purposes. As far as possible, the Z groups in each case were made smaller than the X and Y groups so as to provide for necessary variations in the course of study and to admit of individual child-study and child-diagnosis. In many cases, the reorganization of the school led to special promotions, a topic I shall discuss later. It necessitated also demotions in several instances, but in the main it has been found more politic to place such pupils in special rooms for observation and study than to run the risk of antagonizing teachers and parents at the inception of the new venture.

After the primary reorganization had taken place, the school counselor was asked to devote her time (*a*) to examining new pupils as they entered the school and to place them properly, (*b*) to watching the reorganized groups to permit of frequent readjustments as need might arise, (*c*) to suggesting the formation of one or more special classes—adjustment, development, or opportunity classes—and to assist in their organization, and (*d*) most essential, to studying at least one Z group to find out why the pupils in that group were failing pupils. It seems to us most essential that Z groups should be regarded not as "dumping-grounds" for the educational misfits but rather as educational clinics wherein observation, diagnosis, and remedial measures may be carried on.

Well, these three pioneers went to work intelligently and systematically, handling fellow-teachers and principals with consummate tact. The teachers were made to feel that their judgment received as much if not more consideration than the results of the tests, and that if an obstreperous parent challenged the opinion of the school upon his child the school counselor was right at the teacher's side to give her moral support in the data secured from the tests. When a working classification into groups was arrived at, the results were turned over to the principal, and he, and not the counselor, set the necessary administrative machinery into motion which resulted in the desired reorganization. If a pupil showed plainly by his work that he was misplaced, the counselor was the first person to suggest a new placement in his case. And let me not forget to add that in all this adventure, Dr. Sutherland and his staff gave us the most loyal and whole-hearted coöperation, and thru their larger knowledge and wider experience saved us from many bad blunders.

This was salesmanship of a high order. Soon other principals who had been watching the experiment came in with requests for school counselors, and little by little the pioneer band of three grew slowly until at the present time we have twenty-two persons engaged in the work.

Certain interesting administrative problems arising from the school counselor movement fall to my lot for solution. Some of them were solved, others are still giving us considerable concern. The more experienced counselors came to me soon after the beginning of the present school year, stating that their preliminary placements had been made and that their services were not needed at their several schools for a few weeks until rechecking results would become necessary. We finally worked out a scheme of assignments for such counselors which attaches each to a home school where the counselor must work at stated intervals (i.e. at the beginning, middle, and end of each school term) and which allows them for the balance of the term to take short assignments in other schools which are not large enough to support full-time counselors, but where superficial surveys are necessary and helpful. Obviously, results are not complete as we would like from these short assignments, but we have learned that the school counselor can "sell the idea" at least to a new school and awaken a desire for more extended work in the future.

Another problem arose over special promotions. For example, a child is found in the low fourth grade whose promotion age entitles him to a place in the low fifth grade, an interval of one entire year. The child is promoted and fails to make good, not that his mental ability is not equal to the demands of his new classification but that certain definite steps in the curriculum of the fourth grade have been omitted in his case, and he works under a constant handicap which militates against his success. This problem was solved by placing such a child in an adjustment room either for an hour a day, or in some cases for full-time for a few weeks until the missing steps had been taken and mastered.

The greatest problem of all is, of course, to adapt the technique of instruction in the several groups so as to secure the maximum results. Teachers come to us constantly and say "What shall I do to make my teaching of a Z group successful?" We have instructed the counselors during this spring term to make an intensive study of one Z group—to discover first of all why these pupils are Z pupils. This involves an extensive and intensive case-study of each individual child, and we are finding out some rather extraordinary things. In three of these classes we are trying the experiment of administering modified adjustment materials for two hours a day, and in these groups we have discovered that a number of pupils who have been failures in the fourth grade have failed because they could work only at a high first-grade level in one or more subjects. How much do you suppose a fourth-grade child can profit by repeating the work of the fourth grade when he really needs work at the second-grade level? So the teachers of these three groups have had to grope painfully down and down into the child's experience until they find the initial causes of failure, and then to build him up from that point.

Another problem. I asked the principal of one of the schools where the experiment is being tried what he had gotten out of his study of his Z group, and his answer was that he had been amazed to find that many of the Z pupils were not Z pupils at all. They were, in reality, some

of them Y and X pupils who had been drilled into failure by the blundering administrative machinery of an over-large city school system. Suggest to a child often enough that he is a failure, and after a while the habit of failure becomes fixed and it seems extremely difficult to remove it.

In conclusion, dear friends, let me point out that the whole school counselor movement is merely a means to an end. We are trying to sell to our teachers the idea that modern education means an intensive study of individual differences in children and cannot be expressed in terms of units of subject-matter, pages covered, or lessons taught. We are trying to sell the idea that modern education implies changes in conduct; in the formation of habits, appreciations, and ideals; and only in so far as we recognize these and attempt honestly and sincerely to measure them in quantitative terms are we true to our profession. This necessitates salesmanship of a high order. This necessitates in a true and real sense, educational research.

The Use of Educational Research and Experimentation in School Administration

ELLIS U. GRAFF, *Superintendent of Indianapolis Schools*

THE problem of securing continuity of policy is present in all democracies where public affairs are administered by lay boards. With constant change of personnel this problem is intensified. Public school administration in the United States has undoubtedly been handicapped by frequent changes in boards of education and their executive officers. A recent study appearing in the First Year-Book of the Department of Superintendence shows that the length of service of the superintendent in the entire United States during the period 1900-21 was three years. This period is not long enough for the working out of thoro-going educational policies and obtaining the results inherent in the policies. It has frequently been observed that schools which change their administrative officers and teachers every year do not make much progress. The application of statistical methods to school administration affords an opportunity to obtain a foundation for school policy which can be built upon by successive boards and school executives. This is one of the most important contributions of the statistical method to school administration.

Assuming that the purpose of this discussion is the citing of concrete illustrations of the application of statistical methods to administrative problems rather than a discussion of the theory of statistical methods, allow me to employ as my first example the formulation of a comprehensive building program which was made for our city three or four years ago. At that time the Board of School Commissioners realized that school accommodations were by no means adequate to the needs and that building costs were unusually high, so that it was necessary to solve the problem of additional housing facilities in terms of both efficiency and economy.

The first step in the solution of this problem was the scoring of the 72 existing buildings by the use of the Strayer Score Card for City School Buildings. This card covers all the possible facts with reference to the structure, cost, size, and service systems. It makes possible the scoring of the buildings on a scale of 1,000 points. The result of this survey indicated that the buildings of the city ranked all the way from 241 points out of a possible 1,000 to 803 points, and that 29 buildings scored less than 500 points, or in other words, were less than one-half efficient.

In addition to the scoring, we had an appraisal of all the properties made by a qualified committee, the result of which was the placing of a valuation on every piece of school property. This enabled us to determine whether a given building should be abandoned or repaired and

remodeled. We also made a study of the distribution of persons of school age in the city in order to determine the proper location of contemplated buildings. Tables were also made showing the rate of growth in school population in order to determine the number of rooms necessary to provide for the yearly increase of enrollment. The result of this study showed that the average annual increase in the elementary department for a period of fifteen years was slightly in excess of 1,100 pupils.

In addition to the above facts, we formulated and adopted a set of standardized requirements for elementary school buildings. As a basis for this we used the data furnished by the Committee on Standards of the N.E.A. and supplemented it by the work of our local architects' association and a local committee which visited and studied more than forty modern elementary schools. The result of this study was the adoption of a report containing standardized requirements which insure the obtaining of buildings entirely satisfactory from an educational standpoint. The requirements cover such points as the percentage of areas to be devoted to the various features of school work, the number of rooms of various kinds required, the shape and dimension of the classrooms, and the general layout for the heating, ventilation, and sanitary equipment of the building. Certain items of interior finish are also specified, such as floors, walls, ceilings, wainscot, etc. The adoption of these requirements does not necessitate absolute uniformity of result. The architect is permitted to design the building according to his own ideas so long as he includes in it the requirements contained in the standard specifications. The result has proved very satisfactory. We have obtained buildings which were uniformly satisfactory from an educational standpoint, and yet there has been a certain amount of variety in design and finish. A study of the pamphlet published by the Indianapolis Board of School Commissioners entitled *Standardized Requirements for the Construction of Elementary School Buildings in the City of Indianapolis* will make clear just how both uniformity and variety are obtained.

The four types of information above indicated, that is, building survey, appraisal, study of population, and standardized requirements for buildings, enabled the Board to formulate a building policy covering a period of two years which called for new buildings in various parts of the city with a total capacity of 172 classrooms, and additions to other buildings to the extent of 84 additional classrooms. The report also included new installations of heating and plumbing with estimated costs therefor. The total amount of the building program as then outlined called for \$1,614,398 for the first year's work and \$1,552,855 for the second year's work, or a grand total of \$3,167,253. These estimates were based upon the then current prices of labor and materials. The above items were, of course, restricted to elementary buildings. The high school needs were set forth in a separate study. Part of this projected program was completed within the time specified, but part of it could not be carried out owing to excessive increase in building costs and other difficulties, but within the past six months the Board has again taken up the problem of the building program and we have now em-

ployed architects and are drawing plans for eight additional new elementary buildings and several additions which will be started this season. These new buildings are not exclusively the ones outlined by the first building program, but the need for them was arrived at by reference to the same facts which were used originally. The only modifications were occasioned by new developments in the city, the increase of school enumeration in certain sections due to the building up of suburban communities, and similar facts which could not be foreseen. It is very significant that the type of study indicated furnished facts as the basis of a building program which removed the whole subject from the realm of guess work on the one hand or of petition or public clamor on the other. It has possessed the splendid advantage of satisfying the community that the problem was being solved in the light of the fullest information, and that the ultimate outcome of the whole program would be an adequate and satisfactory building equipment for the schools of Indianapolis.

FINANCIAL PROBLEMS

Another type of study may be of interest for its bearing on the question of school finance. One of the fundamental needs in any organization is a knowledge of costs of maintenance and operation. The modern scientific budget and scientific accounting systems enable us to keep as careful cost finding systems as can be found in any commercial enterprise. A splendid contribution to this department of school administration has been made by the research division of the Intercity Conference which prepared an accounting system which has been adopted by the National Department of Superintendence, the National Association of Business Officials, and the United States Bureau of Education. Accompanying the system is a blank form of reporting which is organized to correspond with the plan of keeping school accounts. This blank, a copy of which I have here, reclassifies and redefines the various types of school activity so that there is no overlapping and no confusion. Heretofore it has been difficult to differentiate between instructional, administrative, and supervisory activity. This report attempts to define all such functions so that the expenditures are charged to the correct items. The gain in this plan is twofold: it enables a school system to know its own situation exactly and also it puts it upon a basis which is exactly comparable with other school systems.

As an example of a cost finding study, permit me to quote from a report recently prepared in regard to one of our own high schools covering the first semester of 1922-23. This study shows that there were 1,984 students enrolled in 119 courses, not including music, physical training, expression, and military training. The average number of courses per pupil was 4.3. The average total enrollment per class was 22.2. The computation of pupil hour cost in the report was based on the number of pupils remaining at the end of the semester as reported by the teacher. A pupil hour of instruction is a sixty-minute period of regularly assigned recitation or laboratory time. Since the class periods are in multiples of forty minutes, a pupil enrolled in a single period

subject for ninety days would require sixty double hours in instruction per semester. The total number of pupil hours of instruction required for all pupils was 501,620, or an average of 65.9 per subject, or 295.9 per pupil. For teachers' salaries alone and considering only the time spent in conducting classes \$94,531.75 was required. The percentage of failures was 12.7, so that approximately \$12,000 worth of instruction did not bring results as far as successful work was measured by school records. The expenditure for administration in this school for this semester was \$11,103.31 or \$5.59 per pupil. For supervision the cost was \$5,673.34 or \$2.86 per pupil. The expenditure for instruction as divided among the various subjects was as follows:

English	23.2 cents per pupil hour
Commercial	14.2 cents per pupil hour
Mathematics	15.5 cents per pupil hour
History	16 cents per pupil hour
Foreign Languages	19.5 cents per pupil hour
Science	19.1 cents per pupil hour
Shop, Mechanical Drawing.....	24 cents per pupil hour
Art	21.1 cents per pupil hour
Domestic Art	18.8 cents per pupil hour

These facts show that the costs vary from 24 cents to 14.2 cents per pupil hour. The cost per pupil hour for the whole school was 18.8 cents. A careful study of these figures enables us to adjust the teaching load fairly to the teaching stabilized to base the class enrollment upon a figure which will be economical. It also prevents a disproportionate amount being expended for supervision or other activities outside of instruction proper. Of course, it goes without saying that cost is not the only nor indeed the major factor to be considered in education, but no doubt all will agree that the costs should be known, and decisions as to the undertaking of various types of activity should be made in the light of a full knowledge of the financial elements involved.

READING STUDY

An example of the value of tests and standards in the teaching of reading will be shown by the following results obtained in our schools by the use of reading tests. As shown by Chart I, the tests given in 1919 revealed the fact that our results in comprehension were slightly below the standard in all grades. Conferences were held in an effort to determine the reasons for the difficulty, and among other considerations was the question of the amount of time devoted to the subject. Chart II was prepared showing the relation of the time spent in reading in the Indianapolis schools to that in fifty cities. You will see from this chart that the time in Indianapolis is considerably less than the average of the fifty cities. Thru repeated conferences and a careful consideration of methods, results were greatly improved, and the tests for the first semester of this year show that our results were above standard in all grades. This was obtained without increasing the time devoted to the subject. Chart III shows a corresponding situation with reference to the speed in reading.

This study was largely diagnostic. It is frequently the case that the correction of a condition is not difficult when the real status of the case is brought to light. Of course, if special adjustments or methods are needed, they can be provided as soon as the exact needs are known.

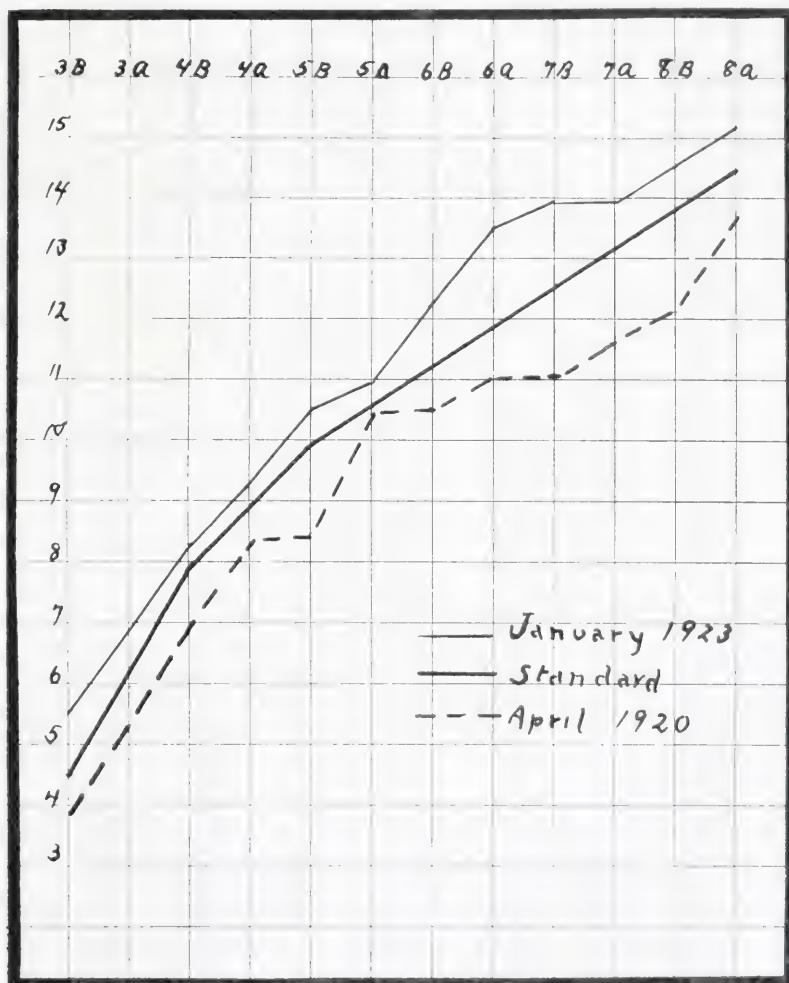


CHART I

Showing the achievements of the Indianapolis Public Schools in comprehension in April, 1920, and in January, 1923, compared with an estimated standard. The tests used were those devised by Monroe to test speed and comprehension in silent readings.

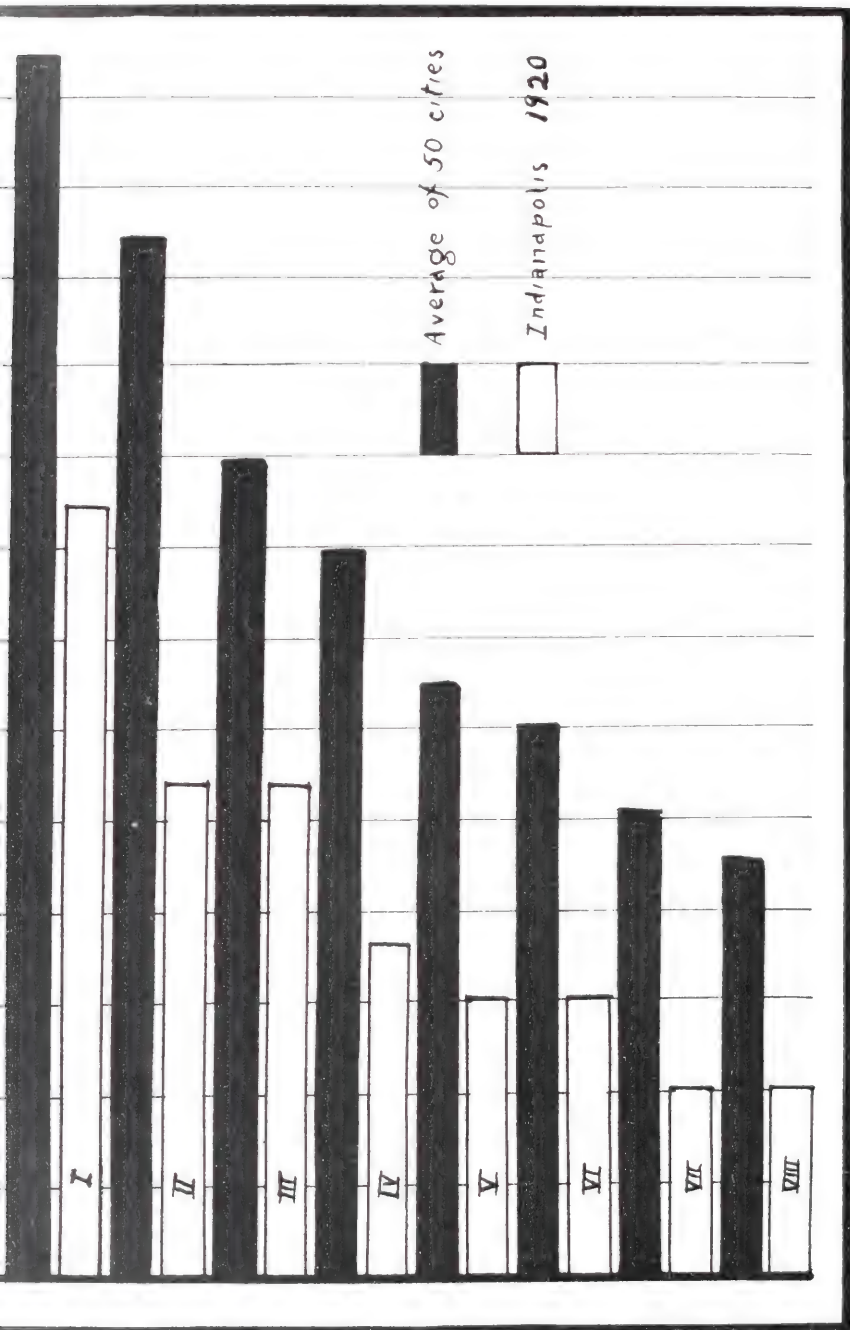


CHART II

Showing the amount of time devoted to reading in each grade in the Indianapolis public schools in 1920 in comparison to the average amount of time devoted to reading in each grade in fifty cities.

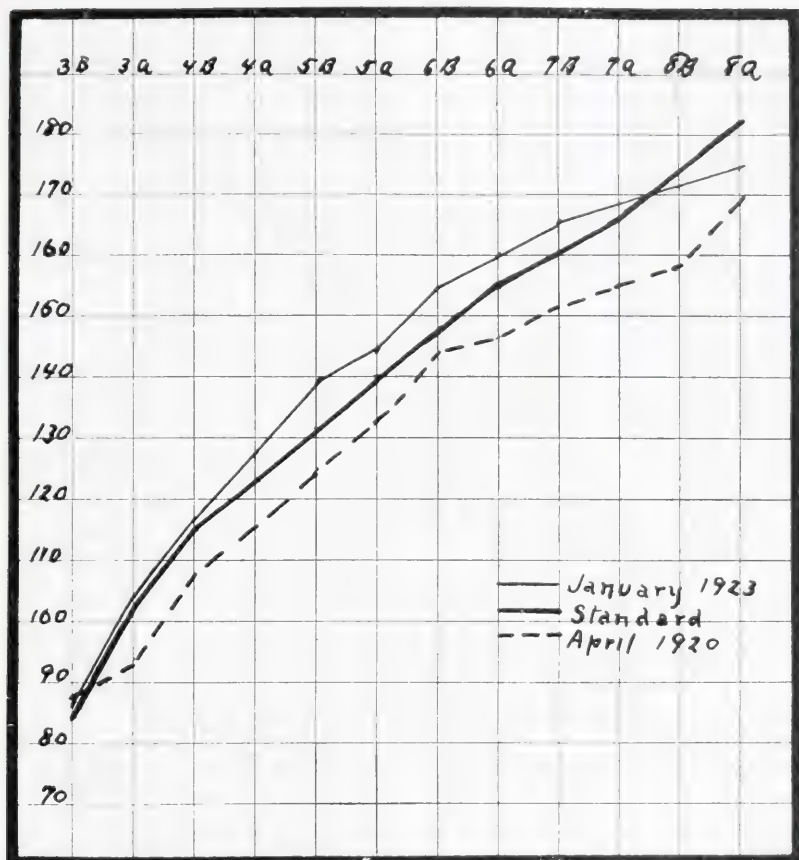


CHART III

Showing achievement of Indianapolis Public Schools in rate of silent reading in April, 1920, and January, 1923, compared with a standard. The tests used were those devised by Monroe to test rate and comprehension in silent reading.

The use of statistical methods in administrative problems is so frequent that it would be impossible in a paper of this sort to cover very much of the field. Very few problems in administration are solved without the use of statistical methods, and the department of research is one of the busiest places in the modern school office. Of the rightness of an opinion of Dr. H. O. Rugg, I think there can be no doubt. In the preface of his book, *Statistical Methods Applied to Education*, he says, "This book is based upon the doctrine that statistical methods in themselves prove nothing,—that the methods selected for use in a particular situation must agree with the logic of that situation: in a word, that statistical methods are merely quantitative devices which we can use to refine our thinking about complex masses of data and to refine our methods of expression."

The Use of Educational Research and Experimentation in School Administration

A. C. SENOUR, *Director of Department of Measurements, East Chicago (Ind.) Schools*

SCHOOL service must commence and terminate with the teacher. What the classroom director does for those under her charge ultimately is the measure of the worth of the school enterprise. That is not questioned. Any discussion of the use in school administration of any program or movement that does not consider its value to or influence upon the teacher necessarily fails of completeness. Perhaps it misses the mark entirely.

In discussing the use of educational research and experimentation in school administration I choose, therefore, as of major importance the value to the teacher. Experimentation is stimulating. It calls for critical analysis of past performances, a weighing of procedures, an alertness as to steps to be taken, an anticipatory state of mind concerning outcomes, and concludes with a feeling of happiness from having attempted the development or creation of something. Experimentation is an antidote for stagnation, a preventive of excessive routinizing.

Last fall in East Chicago an experiment was undertaken to determine the relative value of two different methods of language presentation. A teacher who was ordinarily successful but was not especially enthusiastic in teaching language was selected to conduct the experiment. The chief result of her work may be stated in her words: "I formerly felt the language period a burden. I look forward to it now with pleasure." This was volunteered one day in discussing the experiment with her. Of what inestimable worth to a class is this enthusiasm of the teacher!

The experiment which the University with the coöperation of a number of the teachers thruout the state has recently concluded was similarly productive in East Chicago. Our reading in the intermediate grades is taught departmentally by an especially capable and earnest group of teachers. But even they, and thru them the children, have profited as a result of the increased vigor and stimulation to more searching inquiry which the experiment generated.

Another important use of research to the administrator is the findings and their applications. Educational advance is possible thru theorizing, but that advance is more sure, with less chance for necessity for retreat, if the theorizing is substantiated by actual experimentation. The public schools must rely extensively upon the universities as research centers for progressive ideas of organization and instruction. These institutions have a genuine obligation which they are appreciating more and more. Very many of the recent scientific determinations of

methods of teaching the various subjects are their products. Estimate the value to the administrator of the clean-cut statements for him of the principles of method in instruction in spelling, or in reading, that have been determined largely by university research. How much more definite the course of study constructor can be!

The researches by the educational psychologists in the field of intelligence measurement have provided the administrator with instruments which are gradually culminating in altered organizations. Administrators have recognized for years the wide range in ability of the children grouped into a class for educational purposes. Nevertheless, but little was done until very recently to ameliorate the situation. The administrator's use of the research product, the intelligence tests, to reclassify pupils into more homogeneous groups is a beginning.

At East Chicago classifications into ability groups on the basis of intelligence testing, supplemented by achievement test scores, teachers' rankings, etc., has been begun. All our primer pupils and those entering our junior high school were classified last September and again in February. While we have no objectively obtained data as to the benefits in the first grade, we do have the teachers' recommendations for the continuance of the work on the strength of the increased service they have been enabled to render to those under their charge. In the seventh grade in arithmetic fundamentals, all sections, the strongest to the weakest in ability, made more than the normal gain, the strongest exceeding the normal gain by nearly 100 per cent; the weakest, by approximately 16 per cent. East Chicago contemplates the continuation of this multiple-track organization until all grades are so established.

Classifications, however, produce a new problem—that of curricular adjustments. It is at this stage that the administrator must again employ research and experiment. What differentiations in methods and materials must be made? The answer can be given only from the findings of experimentation. We, at East Chicago, are planning an attack upon the problem. I conceive the situation, tho, as one which might well be undertaken on a more extended scale. Is the problem not one of such coming universal importance that the more progressive schools of the state can coöperatively consider it under the leadership of the University Bureau of Coöperative Research or of the State Department of Education? Let us ponder over this suggestion, not merely assent or dissent.

In conclusion, I desire to mention one of the difficulties encountered in the use of research, one that will never entirely be overcome owing to human nature. It is the false impression which too frequently has been implanted, perhaps not intentionally, by teacher training institutions that methods and devices which are discussed and adopted by teachers in training are the desirable ones and that others must not be considered. The hesitancy on the part of teachers to accept different standards of work from sections or individuals of differing abilities is an illustration. Our training institutions must strive to overcome this attitude in the teachers now in training.

